Projecte de Fi de Carrera

Enginyeria Industrial

Generació d'imatges estereoscòpiques a partir d'una posició qualsevol de les càmeres

MEMÒRIA

Autor: **Bernat García Larrosa**

Director: **Antoni Susín Sánchez**

Convocatòria: **Abril 2011 (Plà 94)**

Escola Tècnica Superior d'Enginyeria Industrial de Barcelona

Resum

Una imatge estereoscòpica és en realitat una parella d'imatges d'una mateixa escena que, mostrades i visualitzades adequadament, donen informació sobre la profunditat dels objectes representats. Actualment aquest tipus d'imatges estan esdevenint cada cop més populars degut a la introducció en el mercat de televisors 3D, i a la cada cop més habitual realització de películes en aquest format.

Per a obtenir una imatge estereoscòpica amb la màxima qualitat, les imatges per cada ull s'han de prendre situant les càmeres a una distància similar a les dels ulls i amb orientacions idèntiques, exàctament igual que en els ulls. Això que en el cos humà s'aconsegueix amb, aparentment, tanta facilitat, reproduir-ho artificialment és realment complicat. Qualsevol petita pertorbació en el muntatge de les càmeres (moviments, dilatacions per canvis de temperatura, etc) pot provocar un allunyament molt gran de les condicions ideals. Per tant, tot i que cada cop s'estan aconseguint sistemes més fiables, encara segueix sent complicat i car.

L'objectiu d'aquest projecte és proposar una alternativa barata per a obtenir imatges estereoscòpiques sense la necessitat de que les càmeres estiguin en les condicions ideals de distància i orientació. S'estudiaran tot un seguit de procediments necessaris per partir de dos imatges preses amb una posició qualsevol de les càmeres, i acabar obtenint una imatge estereoscòpica de l'escena fotografiada.

Índex

Resum **1**

Índex **3**

1 Obtenció d'imatges i rectificació **9**
1.1 Representació d'imatges 9
1.1.1 Imatges en escala de grisos 9
1.1.2 Espais de color 12
1.1.3 Imatges en color 13
1.1.4 Model matemàtic de la càmera fotogràfica 15
1.2 Càmeres i rigs stereo 16
1.2.1 Captació d'imatges estèreo i solucions actuals 17
1.2.2 Dificultats i inconvenients 19
1.3 Geometria epipolar 19
1.4 Detecció i descripció de punts característics 21
1.4.1 Plantejament del problema 21
1.4.2 SURF 22
1.5 Rectificació d'imatges 27

2 Mapes de disparitats **31**
2.1 Concepte de mapa de disparitats 31
2.2 Obtenció del mapa de disparitats 34
2.2.1 Funcions de cost. Mètriques en imatges 34
2.2.2 Implementació en CUDA 37

3 Generació d'imatges basada en mapes de disparitats **41**
3.1 Generació de noves imatges 41

3.1.1 Traslacions . . . 42
3.1.2 Rotacions . . . 43
3.1.3 Moviment general . . . 44
3.2 Ompliment de forats . . . 45
3.2.1 Definició del concepte de forat . . . 45
3.2.2 Mida d'un forat . . . 46
3.2.3 Interpolació . . . 46
3.2.4 Inpainting . . . 48
3.2.5 Interpolació per capes . . . 50
3.2.6 Correcció del mapa de disparitats . . . 51
3.2.7 Comparació entre les diferents alternatives . . . 56
3.3 Resum del procediment utilitzat . . . 56

4 Composició **59**
4.1 Tècniques de composició . . . 59
4.1.1 Anaglif . . . 59
4.1.2 Polarització . . . 61
4.1.3 Ulleres actives . . . 61
4.1.4 Comparació de les diferents tècniques . . . 63
4.1.5 Altres tècniques . . . 63

5 Resultats **65**
5.1 Mores . . . 65
5.1.1 Imatges . . . 65
5.1.2 Temps de càlcul . . . 68
5.2 Art . . . 68
5.2.1 Imatges . . . 68
5.2.2 Temps de càlcul . . . 71
5.3 Oset . . . 71
5.3.1 Imatges . . . 72
5.3.2 Temps de càlcul . . . 74
5.4 Cons . . . 75
5.4.1 Imatges . . . 75
5.4.2 Temps de càlcul . . . 78
5.5 Porta . . . 78
5.5.1 Imatges . . . 79
5.5.2 Temps de càlcul . . . 82

Conclusions **83**

Costos del projecte **85**

Estudi mediambiental **87**

Bibliografia **89**

A Conceptes matemàtics **91**
A.1 Moviments rígids a l'espai 91
A.1.1 Estructura mètrica de l'espai 91
A.1.2 Isometries lineals 94
A.1.3 Moviments rígids afins 96
A.2 Morfologia matemàtica 97
A.2.1 Reticles complets 97
A.2.2 Estructura de reticle complet en les imatges en escala de grisos 100
A.2.3 Erosions i dilatacions 102
A.2.4 Erosions i dilatacions d'imatges en escala de grisos . . 102
A.3 Convolució 104
A.3.1 Definicions i propietats 105
A.3.2 Convolució aplicada al tractament d'imatges 106

Introducció

Les imatges estereoscòpiques han de permetre al cervell poder realitzar el procès de diplopia fisiològica (obtenir informació sobre la profunditat d'un punt a partir de dues imatges).

El procès que se segueix en aquest projecte per tal de generar imatges estereoscòpiques a partir d'una posició qualsevol de les càmeres és:

- Obtenció i representació de les imatges
- Rectificació de les imatges
- Càlcul de les profunditats (mapa de disparitats)
- Generació d'una nova imatge esquerra a la distància apropiada.
- Composició de la imatge estereoscòpica

Capítol 1

Obtenció d'imatges i rectificació

1.1 Representació d'imatges

En aquesta secció s'introduiran els models matemàtics que es faran servir per a les imatges en escala de grisos, els espais de color i els models per a les imatges en color.

1.1.1 Imatges en escala de grisos

Al llarg d'aquest projecte, es consideraran les imatges com a objectes plans (fotografies en paper, pantalles d'ordinador, pintures sobre llenços,...), és a dir que el punt de partida serà una regió $\Omega \subset \mathbf{R}^2$ compacta* i connexa†. A cada punt d'aquesta regió se li assigna un valor de gris, essent el valor mínim el corresponent al negre (menor intensitat de llum) i el valor màxim el corresponent al blanc (màxima intensitat de llum), de manera que el nivell de gris es pot modelar com una aplicació $I : \Omega \subset \mathbf{R}^2 \to [0, \infty) \subset \mathbf{R}$ tal que a cada punt $(x, y) \in \Omega$ li assigna el seu valor de gris $I(x, y)$. Per tant el model matemàtic d'imatge en escala de grisos serà el següent:

*Un subconjunt de $\mathbf{R}^n$ és compacte si és tancat i fitat (Teorema de Heine-Borel)

†De manera poc rigurosa, es pot definir regió connexa com aquella que està formada per una sola peça

Definició 1. *Una imatge en escala de grisos és un parell* (Ω, I)*, sent* $\Omega \subset \boldsymbol{R}^2$ *una regió compacta i connexa, i* $I : \Omega \to [0, \infty)$ *una aplicació que assigna a cada punt* $(x, y) \in \Omega$ *un valor de gris.*

Sovint, si se sobreentén quina és la regió Ω s'anomena imatge en escala de grisos simplement als valors de l'aplicació I. És important remarcar que l'aplicació I no és necessàriament contínua (de fet en la immensa majoria dels casos no ho serà.

Per a poder tractar una imatge amb ordinador, primer de tot cal aplicar-li el procés de *digitalització*, és a dir discretitzar-la per a fer-la apta per a poder ser tractada digitalment. No s'explicarà aquí en què consisteix aquest procés perqué excedeix els objectius d'aquest projecte i d'aquesta memòria, de manera que es passarà a tractar directament el concepte d'imatge digital en escala de grisos. S'ha de dir, això sí, que la necessitat de digitalitzar prové del fet que un ordinador no permet treballar amb regions contínues (en el sentit estricte de continuitat) perqué caldria memòria infinita, i del fet que una pantalla d'ordinador no és contínua sino que és un conjunt de punts, molt propers entre ells per a donar la sensació de continuitat, que emeten llum cadascun d'ells per separat. Així doncs, l'espai de treball passa a ser $\mathbf{Z}^2$ en comptes de $\mathbf{R}^2$, i la regió Ω serà $[1, N] \times [1, M] \subset \mathbf{Z}^2$. A més a més, per les raons exposades, existirà també una limitació en el nombre de valors que pot prendre I, de manera que es prenen 256 nivells de gris, sent 0 el valor assignat al negre i 255 el valor assignat al blanc. Per tant, de manera anàloga a la definició anterior:

Definició 2. *Una imatge digital en escala de grisos és un parell* (Ω, I)*, sent* $\Omega = [1, N] \times [1, M] \subset \boldsymbol{Z}^2$ *amb* $N, M \in \boldsymbol{Z}^+$*, i* $I : \Omega \to [0, 255] \cap \boldsymbol{Z}$ *una aplicació que assigna a cada punt* $(x, y) \in \Omega$ *un valor de gris. Els punts* $(x, y) \in \Omega$ *s'anomenen píxels de la imatge i el nombre* $N \cdot M$ *rep el nom de resolució de la imatge.*

Observacions:

- Igual que abans, sovint se sobreentén la regió Ω i per tant s'identifica la imatge amb els valors que pren I.
- La resolució de la imatge és el nombre de píxels que té aquesta, i dóna una idea del detall amb el que es poden observar els objectes

representats. És, per tant, una mesura de qualitat. Usualment, donat que el valor de la resolució pot ser elevat es dóna en *megapíxels*, així per exemple una imatge amb $N = 1200$ i $M = 1400$ tindrà una resolució de $1200 \cdot 1400 = 1920000 \equiv 1.92$ megapíxels.

- Una imatge digital en escala de grisos es pot pensar com una imatge en escala de grisos considerant la regió $\Omega = [1, N] \times [1, M]$ com un subconjunt de $\mathbf{R}^2$ en comptes de $\mathbf{Z}^2$ i extenent l'aplicació I adequadament.

Exemple 1. *Sigui* $\Omega = [1, 10] \times [1, 10] \subset \boldsymbol{Z}^2$ *i* $I : \Omega \to [0, 255]$ *l'aplicació que ve donada pels següents valors:*

$$\begin{bmatrix} 0 & 14 & 28 & 43 & 57 & 71 & 85 & 99 & 113 & 128 \\ 14 & 28 & 43 & 57 & 71 & 85 & 99 & 113 & 128 & 142 \\ 28 & 43 & 57 & 71 & 85 & 99 & 113 & 127 & 142 & 156 \\ 43 & 57 & 71 & 85 & 99 & 113 & 128 & 142 & 156 & 170 \\ 57 & 71 & 85 & 99 & 113 & 128 & 142 & 156 & 170 & 184 \\ 71 & 85 & 99 & 113 & 128 & 142 & 156 & 170 & 184 & 198 \\ 85 & 99 & 113 & 128 & 142 & 156 & 170 & 184 & 198 & 212 \\ 99 & 113 & 127 & 142 & 156 & 170 & 184 & 198 & 212 & 227 \\ 113 & 128 & 142 & 156 & 170 & 184 & 198 & 212 & 227 & 241 \\ 128 & 142 & 156 & 170 & 184 & 198 & 212 & 227 & 241 & 255 \end{bmatrix}$$

la representació de la imatge digital en escala de grisos corresponent és:

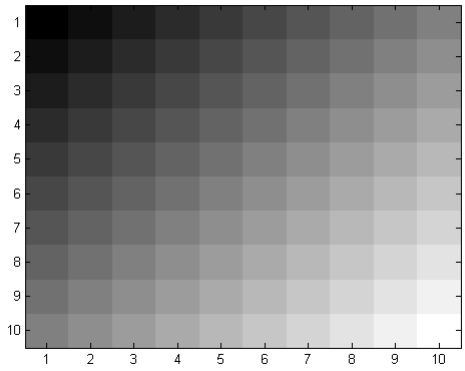


Figura 1.1: Imatge digital en escala de grisos

1.1.2 Espais de color

Per a poder parlar d'imatges en color cal poder parlar abans amb precissió del color en si. L'objectiu d'aquest apartat és, doncs, definir què és el color i definir i explicar què és un espai de color i quins són els espais més habituals.

Intuitivament, la noció de color és molt evident. Donar-ne una definició precisa i rigorosa ja no és tan evident. Segons la Gran Enciclopèdia Catalana, el color és la ***"qualitat de la sensació produïda en un observador per l'efecte distint que provoquen en la retina les llums de diferents longituds d'ona compreses entre uns 380 i uns 760 nm"***, d'aquesta manera quan es diu, per exemple, que un objecte és vermell és perqué emet longituds d'ona compreses entre 620 i 760nm, que són els valors corresponents als colors vermells.

Per a treballar amb colors, ja sigui en visió artificial, art, televisió, fotografia... es necessita establir un model de color, i això és justament el que proporcionen els espais de color. De manera poc rigorosa es pot dir que els colors es poden combinar linealment entre ells o, dit d'una altra forma, mesclar-los, per exemple agafant pots de pintura de colors diferents i barrejant unes determinades quantitats de cada pot per a obtenir una pintura d'un nou color.

Definició 3. *Un espai de color de dimensió n és el conjunt format per totes les combinacions lineals de n colors independents (colors tals que cap d'ells es pot obtenir com a combinació dels altres).*

Per exemple, l'escala de grisos seria un espai de color de dimensió 1 prenent com a base el color blanc. Els espais de color més habituals tenen dimensió 3: RGB, RYB... Donats dos espais de color, es dirà que un és millor si permet obtenir com a mínim els mateixos colors que l'altre.

La intenció és determinar un espai de color de dimensió baixa i que sigui el millor posible. Cal diferenciar dos situacions: que l'objecte que es vol tractar emeti el color per reflexió de llum (pintures, papers, fotografies,...), o que el propi objecte sigui una font emisora de llum (pantalles, leds, lletrers lluminosos...). En el primer cas la formació de colors és mitjançant el que es coneix com a síntesis substractiva, i en el segon cas com a síntesis additiva. La raó d'aquests noms és perque, per exemple, si es mescla un pot de pintura

vermella i un pot de pintura blava, experimentalment es pot comprovar com el color resultant serà negre, és a dir que mesclant una pintura que reflecteix el color vermell amb una que reflecteix el color blau, el resultat és una nova pintura que no reflecteix cap color, d'aquí el nom de síntesis substractiva. En canvi en la síntesis additiva, com que la font emisora de llum és el propi objecte, a mesura que se superposen objectes s'aniran emetent cada cop més longituds d'ona (tantes com objectes amb longituds d'ona diferents s'ajuntin).

En el cas de la síntesis substractiva, un primer intent és l'espai RYB (Red, Yellow, Blue), que pren com a base els colors vermell, groc i blau. Amb el temps es va acabar demostrant que aquest espai de color no era bo, en el sentit de que no permet aproximar bé tots els colors. Una millora d'aquest és l'espai CMYK (Cyan, Magenta, Yellow, Key), espai de dimensió 4 que pren com a base el cian, el magenta, el groc i el negre. Aquest espai de color és el que es fa servir habitualment en la impresió de colors.

En el cas de la síntesis additiva, l'estratègia més inteligent és fixar-se en l'anatomia de l'ull. Les cèlules encarregades de captar la llum són els cons i els bastons. D'aquestes, les que capten el color són els cons, i d'aquests hi ha tres tipus: els que capten llum blava, els que capten llum vermella i els que capten llum verda. Sembla, per tant, que la opció més raonable és prendre com a base aquests tres colors, i això és el que dona lloc a l'espai RGB. Aquest espai és l'usat en pantalles i monitors, i serà per tant el que s'usarà en aquest projecte. Així doncs, d'ara en endavant, llevat que s'especifiqui el contrari, se suposarà que es treballa amb l'espai de color RGB.

1.1.3 Imatges en color

La idea per a poder parlar d'imatges de color és molt similar a la manera de modelar les imatges en escala de gris. Es pot pensar en cadascun dels tres canals de color per separat, vermell, verd i blau, i assignar a cada punt de la imatge un valor d'intensitat de vermell, un de verd i un de blau, exactament igual que s'assigna un valor de gris en les imatges en escala de grisos.

Definició 4. *Una imatge en color és un parell* (Ω, I)*, sent* $\Omega \subset \boldsymbol{R}^2$ *una regió compacta i connexa, i* $I : \Omega \to [0, \infty)^3$ *una aplicació que assigna a*

cada punt $(x, y) \in \Omega$ *un valor de vermell, un valor de verd i un valor de blau.*

I de la mateixa manera que amb les imatges en escala de grisos, es pot definir el concepte d'imatge digital en color:

Definició 5. *Una imatge digital en color és un parell* (Ω, I)*, sent* $\Omega = [1, N] \times [1, M] \subset \boldsymbol{Z}^2$ *amb* $N, M \in \boldsymbol{Z}^+$*, i* $I : \Omega \to [0, 255]^3 \cap \boldsymbol{Z}$ *una aplicació que assigna a cada punt* $(x, y) \in \Omega$ *un valor de vermell, un valor de verd i un valor de blau. Els punts* $(x, y) \in \Omega$ *s'anomenen píxels de la imatge i el nombre* $N \cdot M$ *rep el nom de resolució de la imatge.*

Sovint serà útil escriure $I = (I_R, I_G, I_B)$, on cada component de I és la intensitat d'un color (vermell, verd i blau respectivament). Igual que en el cas de l'escala de grisos, es poden fer les següents observacions:

- Se sobreentén la regió Ω i per tant s'identifica la imatge amb els valors que pren I.
- Una imatge digital en color es pot pensar com una imatge en color considerant la regió $\Omega = [1, N] \times [1, M]$ com un subconjunt de $\mathbf{R}^2$ en comptes de $\mathbf{Z}^2$ i extenent l'aplicació I adequadament.

Exemple 2. *Sigui* $\Omega = [1, 10] \times [1, 10] \subset \boldsymbol{R}^2$*, i la següent imatge digital en color: Alguns valors d'I són:*

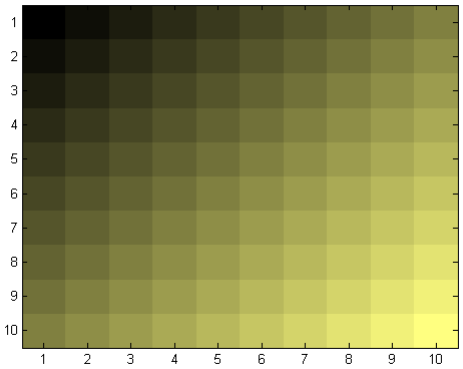


Figura 1.2: Imatge digital en color

$$I(1,1) = (0,0,0)$$

$$I(10,1) = (128,128,64)$$

$$I(1,10) = (128,128,64)$$

$$I(5,5) = (113,113,57)$$

$$(10,10) = (255,255,128)$$

1.1.4 Model matemàtic de la càmera fotogràfica

Donat un punt de l'espai del qual se'n fa una fotografia, el model matemàtic de la càmera fotogràfica permet establir quina és la relació entre les coordenades d'aquest punt i el píxel que el representa en la imatge que s'obtindrà. El model en qüestió s'ha extret de Ma [14].

S'assumirà el compliment de les següents hipòtesis:

- La distància focal de la lent de la càmera tindrà un valor conegut, f.
- L'origen de coordenades se situarà al centre de la càmera.
- Es prendrà com a eix Z la direcció perpendicular a la lent de la càmera.

A més a més, es suposarà que la lent de la càmera és infinitament petita, és a dir que ocupa tant sols un punt de l'espai. Aquesta suposició dona lloc a l'anomenat model pinhole. L'acció de fer una fotografia consisteix en projectar els punts d'una escena sobre un pla anomenat pla de la imatge.

Definició 6. *S'anomena pla de la imatge al pla paralel al pla càmera que es troba a distància f d'aquest.*

El sistema de coordenades que es pren al pla de la imatge consisteix en agafar com a origen la projecció ortogonal del centre de la càmera sobre el pla de la imatge, és a dir el punt de coordenades espaials $(0,0,f)$ (ja que l'origen de coordenades espaials s'ha situat al centre de la càmera), i com a base els vectors $(1,0,0)$ i $(0,1,0)$, de manera que es pot identificar el pla de la imatge amb $\mathbf{R}^2$ i el subconjunt Ω del pla de la imatge on es projectaran els punts de l'escena es pot pensar com un subconjunt de $\mathbf{R}^2$.

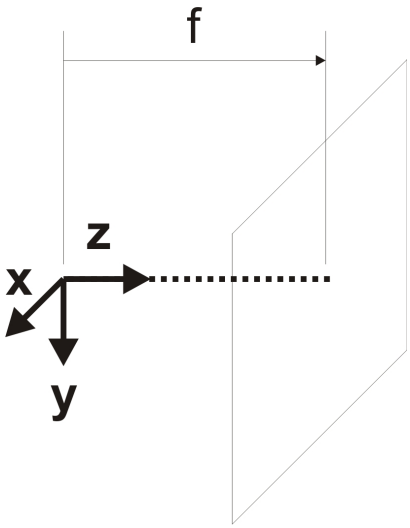


Figura 1.3: Pla de la imatge

L'aplicació que envia un punt $P = (X, Y, Z)$ a la seva projecció s'anomenarà π, i la notació que es farà servir serà $(x, y) = \pi(X, Y, Z)$. A la matriu de l'aplicació π en coordenades projectives se l'anomena matriu de la càmera, i es denotarà per R. En la base canònica,

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = R \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Com que en aquest projecte es treballarà bàsicament amb imatges digitals en color, se suposarà que les imatges obtingues amb la càmera són d'aquest tipus. Així doncs, el pla de la imatge serà realment $\mathbf{Z}^2$, i quan es projecti un punt $P = (X, Y, Z)$ caldrà arrodonir les coordenades de la projecció per a que siguin nombres enters.

1.2 Càmeres i rigs stereo

En aquest apartat s'explicarà breument la tècnica de captació d'imatges estereoscòpiques, es mostraran algunes de les solucions que es fan servir actualment i es veuran les dificultats i els inconvenients que poden sorgir.

1.2.1 Captació d'imatges estèreo i solucions actuals

Com ja s'ha comentat a la introducció, per a generar una imatge estèreo cal prendre dues imatges d'una mateixa escena, una per a ser visualitzada amb l'ull esquerra i una altra per l'ull dret, per a proporcionar al cervell la mateixa informació que rebria si l'escena fos observada a la realitat. Però justament per aquest motiu, cal aconseguir captar les imatges de la manera més semblant possible a com ho farien els ulls, és a dir, situant les càmeres a la mateixa alçada, orientant-les de manera idèntica, i posant-les a una distància apropiada. Les càmeres han de ser idèntiques o, al menys, tenir la mateixa distància focal.

Per a fer això, o bé es construeix un rig estèreo, consistent en muntar les dues càmeres sobre un suport que les mantingui fixes i correctament orientades, o bé es construeix una càmera amb dos lents que capta les dues imatges.

A continuació es mostren alguns exemples de rigs i càmeres estèreo:

Figura 1.4: Rig estèreo amb dos càmeres Panasonic HDC-SD9 CCD

Font: http://www.ssontech.com

Figura 1.5: Rig estèreo amb dos càmeres Canon Powershot

Font: http://www.andrewhazelden.com

Figura 1.6: Càmera estèreo Fujifilm FinePix Real 3D

Font: http://fujifilm.co.uk

Figura 1.7: Càmera analògica Kodak Stereo Camera de l'any 1954

Font: http://en.wikipedia.org

1.2.2 Dificultats i inconvenients

El principal inconvenient dels rigs estèreo és la dificultat per a aconseguir la rigidesa que es requereix. Cal que, en tot moment, les dues càmeres conservin una orientació idèntica, que estiguin sempre a la mateixa alçada i a la mateixa distància relativa entre elles. Això implica una selecció de materials apropiada, per evitar deformacions amb el temps i dilatacions per l'augment de temperatura, i un procés de fabricació molt acurat.

En quant a les càmeres estèreo, justament pels mateixos requeriments de rigidesa en quant a la distància i orientació de les lents, el procés de fabricació és més car.

1.3 Geometria epipolar

Abans de poder parlar de la detecció de punts característics i la rectificació de les imatges, cal saber quina informació serà rellevant per a poder realitzar aquests procesos. La geometria epipolar serveix per a justificar la forma com es planteja la rectificació d'imatges.

Definició 7. *Donades dues càmeres amb la mateixa distància focal, situades als punts O_l i O_r (centres de la càmera esquerra i dreta respectivament), i*

un punt P no alineat amb O_l i O_r, s'anomena pla epipolar de P al pla que conté els punts O_l, O_r i P.

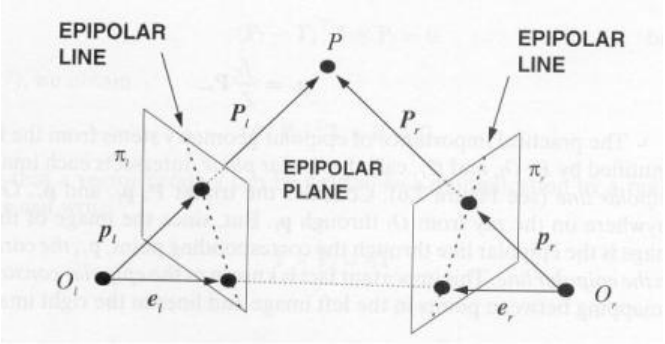


Figura 1.8: Pla epipolar de P

Definició 8. *En les mateixes condicions de la definició anterior, la intersecció del pla epipolar amb els plans de les imatges està formada per les línies epipolars de P, i la intersecció de la recta que uneix els centres de les càmeres amb els plans de les imatges és el parell de punts anomenats epipols.*

Observació: Els epipols no depenen del punt P

En el cas ideal per a poder obtenir dues imatges adequades per a composar una imatge estèreo, és a dir amb les càmeres amb la mateixa orientació, les línies epipolars d'un punt P seran les mateixes als dos plans de la imatge, i els epipols no existiran, ja que la recta d'unió dels centres de les càmeres serà paralela als plans de les imatges (en termes de geometria projectiva es dirà que els epipols es troben a l'infinit).

S'observa a més que prenent un sistema de referència tal que la línia que uneix ambdues càmeres sigui horitzontal, aleshores les línies epipolars seran també horitzontals. Per tant, la situació desitjada serà aquesta. És a dir, el procés de rectificació consistirà en aplicar transformacions afins a les imatges per a aconseguir que les línies epipolars de cada punt P de l'escena siguin horitzontals i estiguin a la mateixa alçada.

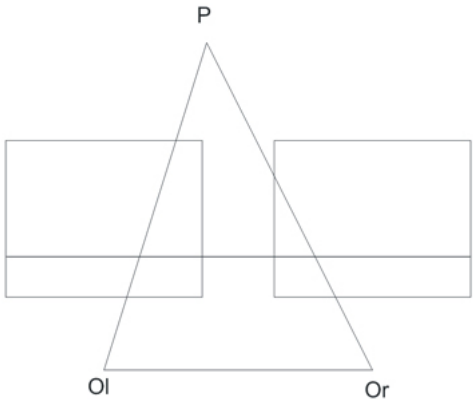


Figura 1.9: Geometria epipolar amb la mateixa orientació de les càmeres

1.4 Detecció i descripció de punts característics

1.4.1 Plantejament del problema

Per tal de convertir en horitzontals les línies epipolars, abans cal conèixer alguns punts equivalents en les dues imatges per poder determinar així quina transformació caldrà aplicar. Per a determinar punts equivalents caldrà que aquests siguin suficientment característics i fàcilment identificables a ambdues imatges. Per exemple, en la parella d'imatges de la figura 1.10 els punts assenyalats amb color blau no són interessants ja que es poden confondre fàcilment amb altres colors del cel. No obstant els punts vermells són fàcilment identificables en ambdues imatges. Així doncs, la recerca de punts característics s'haurà de centrar en vèrtexos, cantonades, punts amb forma de T,... És a dir, punts d'una escena que siguin fàcilment identificables sota qualsevol punt de vista. Tradicionalment, per detectar punts característics s'han fet servir detectors de cantonades com el detector de Harris [2], l'algoritme de Shi i Tomasi, SUSAN (Smallest Univalue Segment Assimilating Nucleus), detectors de canvis d'iluminació com la diferència de gaussianes... Un cop s'han detectat els punts característics, s'assigna a cadascun d'ells un vector anomenat descriptor que emmagatzema informació sobre el punt en qüestió, i aleshores per a identificar un punt d'una imatge amb el seu punt corresponent a l'altra imatge cal definir i minimitzar una distància entre els vectors descriptors (distància euclidiana, distància

Figura 1.10: Punts característics

de Mahalanobis,...). El problema d'aquests algoritmes i dels descriptors que s'han fet servir clàssicament és que no tots són robustos o invariants respecte de canvis i pertorbacions en les imatges. Per aquest motiu es van desenvolupar algoritmes com el SIFT (Scale Invariant Feature Transform), explicat a [9], que proporciona uns punts característics i uns descriptors invariants a canvis d'escala, rotacions, i robustos per canvis d'iluminació, soroll,... i més recentment el SURF (Speed-Up Robust Features), [4], que proporciona resultats similars al SIFT però amb una velocitat de càlcul molt més elevada.

1.4.2 SURF

En aquest apartat s'explicarà el mètode emprat en aquest projecte per a la detecció i descripció de punts característics, el mètode SURF. Primer s'explicarà l'algoritme per a detectar punts i seguidament s'introduirà el vector de descriptors que es farà servir. Com que aquest algoritme s'agafarà ja implementat, no s'aprofundirà molt en detalls tècnics referents a la implementació.

Detecció de punts característics

Abans de començar la detecció de punts característics caldrà fer un filtratge gaussià de la imatge (concretament es farà un filtre per la laplaciana d'una

gaussiana) per tal d'atenuar el possible soroll que hi pugui haver.

Com s'ha dit, interessa que els punts característics siguin invariants a canvis d'escala, per tant la idea inicial consisteix en anar reduint la mida de la imatge a diferents escales i prendre com a punts característics els que siguin màxims locals del determinant de la matriu Hessiana $\mathcal{H}((x,y),\sigma)$,

$$\mathcal{H}((x,y)\sigma) = \begin{pmatrix} L_{xx}((x,y),\sigma) & L_{xy}((x,y),\sigma) \\ L_{xy}((x,y),\sigma) & L_{yy}((x,y),\sigma) \end{pmatrix}$$

on L_{xx}, L_{yy}, L_{xy} representen el filtrat per la laplaciana de la guassiana.

Per determinar si un píxel correspon a un màxim es compara amb els seus veins, considerant el veinat format per les finestres 3×3 en la imatge on s'està cercant el màxim i les seves veines en la representació d'escala (figura 1.11).

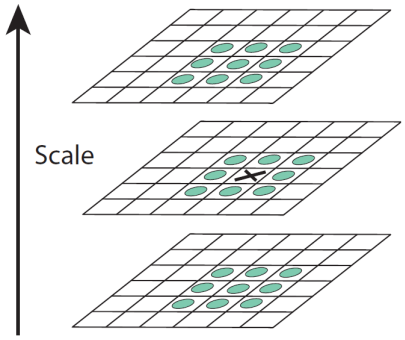


Figura 1.11: Detecció dels màxims de la matriu Hessiana prenent un veinat de 26 píxels format per una finestra 3×3 al voltant del píxel estudiat i finestres 3×3 en les escales adjacents.

Font: Lowe [9]

En comptes de procedir així, però, el que es farà serà mantenir la mida de la imatge constant i variar els valors de la dispersió de la gaussiana, σ. L'avantatge de fer-ho així és que s'evitarà el fenomen d'aliasing i que els càlculs podran ser implementats de forma més eficient, ja que sempre es realitzaran sobre la mateixa imatge original (poden fins i tot paralelitzar-se, per exemple amb CUDA).

A la pràctica, però, no poden aplicar-se els filtres reals, ja que s'estarien aplicant filtres continus a un senyal discret. Cal, doncs, discretitzar i truncar

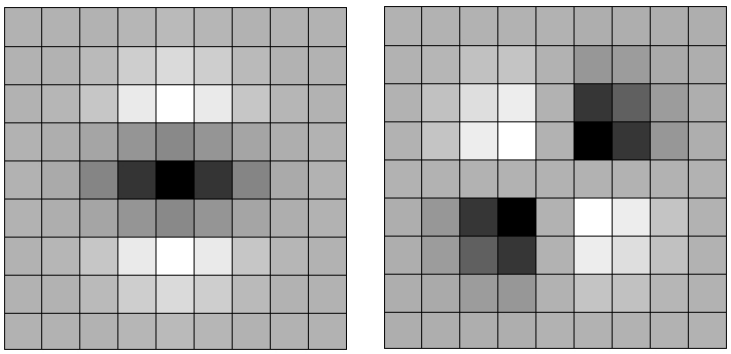


Figura 1.12: Laplacianes de gaussianes reals, L_{yy}, L_{xy}

Font: Bay [4]

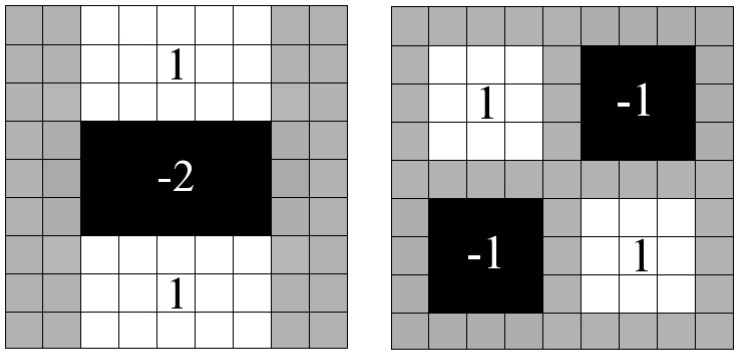


Figura 1.13: Laplacianes de gaussianes discretitzades i truncades, D_{yy}, D_{xy}

Font: Bay [4]

els filtres. Els filtres discrets s'anomenaran D_{xx}, D_{yy}, D_{xy}. Com es pot apreciar en la figura 1.14, la mínima diferència en les mides de dos filtres d'escala consecutiva per tal de preservar les simetries ha de ser de 6 píxels. Per tant, el menor filtre possible serà el de mida 9×9, que correspon aproximadament a una escala $\sigma = 1.2$. Les escales dels següents filtres seran $\sigma = 1.2(6k/9 + 1)$, per $k \geq 1$. Aleshores, la funció a maximitzar serà

$$F((x, y), \sigma) = \det(\mathcal{H}_{aprox}((x, y)\sigma)) = D_{xx}D_{yy} - (wD_{xy})^2$$

on w és un pes que es tria adequadament per tal de que es conservi la energia entre la gaussiana original i la gaussiana discretitzada. En el cas

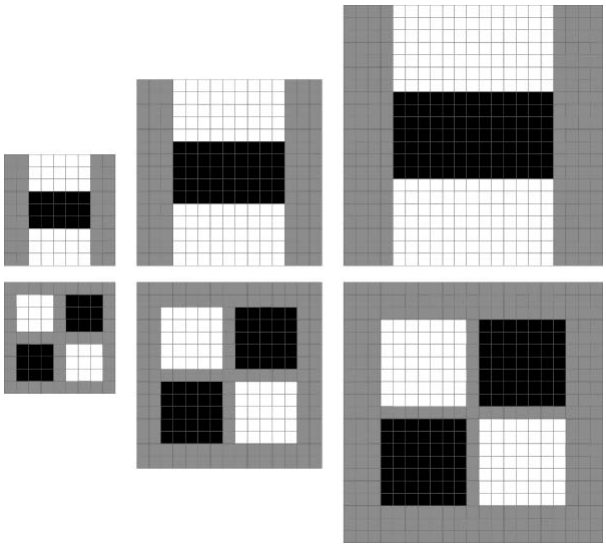


Figura 1.14: Filtres de diferents mides

Font: http://webscript.princeton.edu/ kjenkins/

del filtre més petit,

$$w = \frac{\|L_{xy}(1.2)\|_F \|D_{yy}(9)\|_F}{\|L_{yy}(1.2)\|_F \|D_{xy}(9)\|_F} \approx 0.912$$

on $\|\cdot\|_F$ és la norma de Frobenius. Tot i que, teòricament, el valor de w varia amb l'escala, en l'algoritme SURF es manté constant ja que, segons els autors de Bay [4], els resultats no es veuen gaire alterats.

Descripció de punts característics

A cada punt característic se li assignarà un vector descriptor de 64 components. Per tant, donades dues imatges d'una mateixa escena, dos punts característics seran homòlegs quan la distància entre els seus descriptors sigui mínima.

El primer pas és l'assignació d'una orientació a cada punt. Per a fer això, es calcula la resposta dels wavelets de Haar en les direccions x i y en un veinat circular de radi $6s$ centrat en el punt d'interès, essent s l'escala a la que el punt en qüestió ha estat detectat. Per calcular la resposta dels wavelets de Haar, es fa la convolució amb nuclis formats pels valors -1 i 1, i la mida dels nuclis és $4s$, on s és l'escala (figura 1.15). Seguidament, es filtren

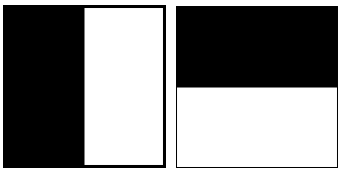


Figura 1.15: Filtres per calcular la resposta dels wavelets de Haar en les direccions x (esquerra) i y (dreta)

Font: Bay [4]

les respostes dels wavelets amb gaussianes de dispersió $\sigma = 2s$ centrada al punt d'interés, i es representen les respostes en un pla, assignant al valor de l'abscissa la resposta en la direcció x i en l'ordenada la resposta en la direcció y. Aleshores, per calcular la orientació del punt d'interès, s'escombra el pla amb un sector circular d'amplada $\pi/3$, calculant les sumes de les components x i y de cada resposta continguda en el sector, formant un vector candidat a ser el vector d'orientació. Aleshores, la direcció s'assignarà a la que té el vector candidat més llarg (figura 1.16). Un cop està assignada l'orientació,

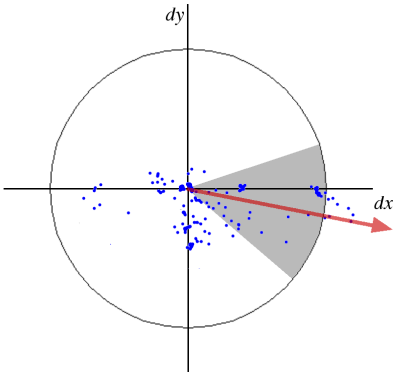


Figura 1.16: Assignació de la orientació

Font: Bay [4]

se centra en cada punt d'interés un quadrat de mida $20s$ orientat amb la orientació assignada al punt en qüestió (figura 1.17) Aquest quadrats se subdivideixen en 16 regions quadrades iguals i dins d'aquests quadrats es pren una graella de 5×5 punts equiespaiats, per a cadascun dels quals es calcula la resposta del wavelet de Haar, essent la direcció horitzontal la corresponent a la direcció del punt, i la vertical la seva perpendicular en sentit positiu. Les respostes horitzontal i vertical s'anomenaran dx i dy respectivament. A la pràctica es calcula en les direccions horitzontal i

Figura 1.17: Punts d'interès amb els seus quadrats corresponents orientats segons la orientació del punt

Font: Bay [4]

vertical sense rotar, i dx, dy es calculen per interpolació. Aleshores, per a cada subregió es forma un vector consistent en sumar totes les respostes horitzontals i verticals, i els seus valors absoluts:

$$\mathbf{v} = \left(\sum dx, \sum dy, \sum |dx|, \sum |dy|\right)$$

I, finalment, el descriptor del punt es construeix concatenant els vectors de cada subregió, obtenint doncs un vector de 64 components.

1.5 Rectificació d'imatges

Un cop es tenen identificats els parells de punts equivalents en les dues imatges, el següent procediment és la rectificació: aplicar una transformació a les imatges per tal que les línies epipolars esdevinguin horitzontals, i que les línies epipolars equivalents estiguin a la mateixa alçada a les dues imatges.

Per identificar la línia epipolar corresponent a un píxel m, tant sols cal fer la intersecció del pla de la imatge amb el pla engendrat pel punt m i els centres de les càmeres (figura 1.18).

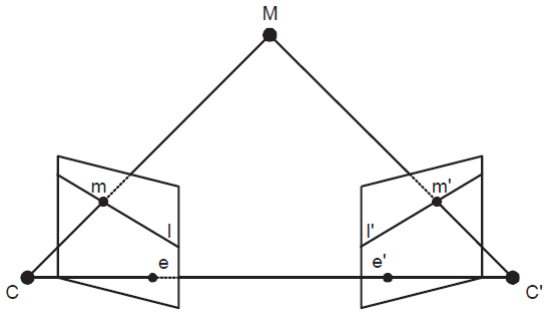


Figura 1.18: La línia epipolar l és la intersecció del pla de la imatge amb el pla que conté els punts C, C' i m (anàlogament per l'). *Observació: no cal conèixer el punt original M, amb conèixer els punts m i m' n'hi ha prou*

Font: Oram [12]

La idea de la rectificació és partir de la les matrius de les càmeres esquerra i dreta, P_{ol}, P_{or}, i aplicar-lis una rotació per posar les càmeres amb la mateixa orientació, garantint així que les línies epipolars siguin horitzontals, és a dir, el que en termes de geometria projectiva es diu que els epipols són a l'infinit. Finalment, si s'escau, caldria aplicar una traslació vertical a una de les dues imatges per posar les línies epipolars a la mateixa alçada.

La situació ideal seria tenir el sistema calibrat, és a dir, conèixer les matrius de les càmeres, però això implicaria poder mesurar o conèixer la posició i orientació de les càmeres, i en aquest projecte es parteix de que les càmeres estan en una posició qualsevol i que aquesta no és coneguda. Per a calibrar imatges en aquestes condicions, se segueix el procediment proposat al paper de Fusiello [1]. En aquest mètode la única informació que cal conèixer són algunes parelles de punts corresponents en la imatge esquerra i dreta, i aquesta és justament la informació proporcionada pel SURF.

Suposant, doncs, que es disposa d'un cert nombre de parelles de punts corresponents, $\mathbf{m}_l^j$, $\mathbf{m}_r^j$, es tracta d'aplicar transformacions a les imatges per tal que cada parella de punts corresponents estigui sobre una línia epipolar i que aquesta sigui horitzontal. Les transformacions que cal aplicar són òbviament desconegudes. La transformació per la imatge dreta s'anomenarà H_r, i per l'esquerra H_l; per tant, al rectificar les imatges els punts $\mathbf{m}_l^j$ i $\mathbf{m}_r^j$

passaran a ser $H_l\mathbf{m}_l^j$ i $H_r\mathbf{m}_r^j$. Així doncs, la condició d'epipolaritat és

$$(H_r\mathbf{m}_r^j)^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} (H_l\mathbf{m}_l^j) = 0$$

Definint la matriu fonamental com

$$F := H_r^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} H_l$$

l'equació anterior es reescriu com:

$$(\mathbf{m}_r^j)^t F \mathbf{m}_l^j = 0$$

Cal, doncs, imposar aquesta equació per a totes les parelles de punts relacionats, i resoldre el sistema per obtenir els coeficients de H_l i H_r. A la realitat, però, aquesta condició serà impossible de satisfer simultàniament per a totes les parelles de punts, degut a errors numèrics i errors i imprecisions en el posicionament dels punts característics. Per tant, plantejant l'equació anterior per a totes les parelles de punts característics relacionats, $(\mathbf{m}_l^j, \mathbf{m}_r^j)$, s'obtè un sistema d'equacions sobredeterminat que caldrà resoldre mitjançant mínims quadrats. No obstant, a Fusiello [1] es proposa una expresió alternativa amb més sentit geomètric, coneguda com a error de Sampson:

$$E_S^j = \frac{((\mathbf{m}_r^j)^t F \mathbf{m}_l^j)^2}{\left\|AF\mathbf{m}_l^j\right\|^2 + \left\|(\mathbf{m}_r^j)^t FA\right\|^2}$$

on

$$A = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

i el sistema sobredeterminat a resoldre és

$$E_S^j = 0, \quad \forall j$$

que en aquest cas és no lineal. A Fusiello [1] es proposa fer servir l'algoritme de Levenberg-Marquardt (funció lsqnonlin de Matlab). Un cop resolt el sistema, i coneguts per tant els coeficients de H_l i H_r, ja poden rectificar-se les imatges.

Capítol 2

Mapes de disparitats

Per a determinar la profunditat d'un punt d'una escena, el procediment més important que fa el cervell és processar les imatges capturades per cada ull per tal d'identificar el punt en qüestió en ambdues imatges. Aleshores, mesurant la diferència en la posició del mateix punt en les dues imatges s'obté una mesura indirecta de la profunditat. Contra més separats estiguin els punts, més proper serà l'objecte i viceversa. Aquest procediment s'anomena diplopia fisiològica. Tot i que també es fan altres processos, com obtenir informació de la profunditat a partir de l'il·luminació i les ombres, del moviment i la mida dels objectes, etc. Per exemple, si una persona veu un vaixell amb una mida molt petita sabrà que aquest està lluny (figura 2.1).

En aquest capítol s'explicarà com reproduir de manera artificial la diplopia fisiològica, emmagatzemant la informació en els anomenats mapes de disparitats.

2.1 Concepte de mapa de disparitats

Donada una imatge digital (en escala de grisos o en color), de manera informal es pot dir que un mapa de disparitats és una imatge digital en escala de grisos que dóna informació sobre la profunditat a l'espai dels punts re-

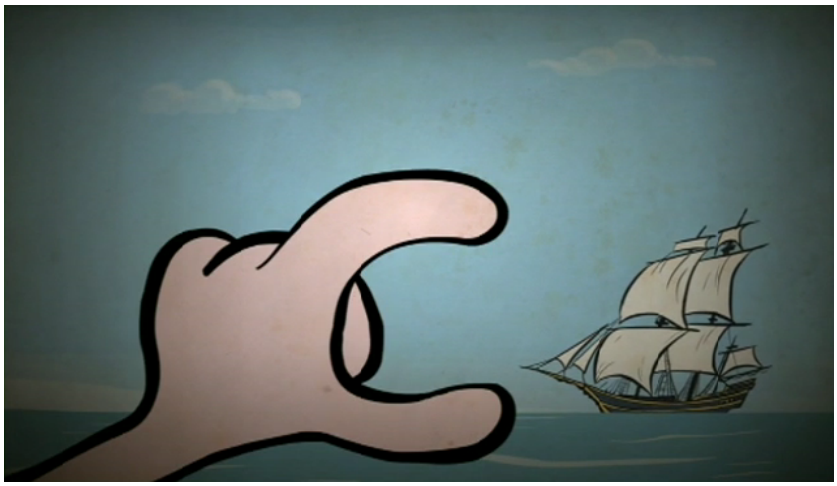


Figura 2.1: Determinació de la profunditat a partir de la mida d'un objecte. Font: Captura dels dibuixos animats *Maricón y Tontico* (Antena3-Neox)

presentats a cada píxel.

Per a crear un mapa de disparitats es necessiten 2 imatges diferents de la mateixa escena, la situació ideal seria disposar de dues imatges preses amb dues càmeres iguals (de fet amb que tinguin la mateixa distància focal n'hi ha prou) que estiguin a la mateixa alçada i amb la mateixa orientació, és a dir que tant sols pot haver-hi una traslació horitzontal entre una càmera i l'altra. Però com ja s'ha comentat, en aquest projecte l'objectiu no és prendre imatges en aquestes condicions tan rígides, sino prendre les imatges amb una posició qualsevol de les càmeres i rectificar-les després. Així doncs, se suposarà que les imatges ja estan rectificades. Se suposaran coneguts els següents paràmetres:

- Distància entre els centres de les imatges (baseline): B (en mm)
- Distància focal de les càmeres: f (en píxels)
- Coordenades espaials de cada punt representat a la imatge (en mm)

Segons s'ha vist anteriorment, segons el model matemàtic de la càmera fotogràfica, la projecció d'un punt $P = (X, Y, Z)$ sobre el pla de la imatge serà (en coordenades del pla de la imatge):

$$(x, y) = \left(f\frac{X}{Z}, f\frac{Y}{Z} \right)$$

Traslladant la càmera segons el vector $(B, 0, 0)$, la projecció del punt $P = (X, Y, Z)$ sobre el nou pla de la imatge ara serà:

$$(x', y') = \left(f\frac{X - B}{Z}, f\frac{Y}{Z} \right)$$

La disparitat del píxel (x, y) és el desplaçament que experimenta aquest al desplaçar la càmera:

$$\text{disp}_B(x, y) = |x - x'| = f\frac{|B|}{Z}$$

El subíndex B serveix per a remarcar que la disparitat depèn de la distància entre càmeres. De totes formes, d'ara en endavant s'ometrà aquesta dependència i s'escriurà simplement $\text{disp}(x, y)$. Com es pot observar, coneguda la disparitat d'un píxel es pot recuperar fàcilment la profunditat del punt que representa, $Z = f\frac{|B|}{\text{disp}(x,y)}$, i això fa que a vegades els mapes de disparitats s'anomenin també mapes de profunditats.

Per tal de poder codificar les disparitats en una imatge digital en escala de grisos, es faran les següents suposicions:

- En cas de no poder-se calcular la disparitat d'un píxel (x, y) s'assignarà el valor $\text{disp}(x, y) = 0$

- La disparitat màxima serà de 255. Aquesta suposició és raonable, ja que per a calcular mapes de disparitats la distància entre les càmeres, B, no serà excessivament gran com per a provocar grans disparitats.

Així doncs, el mapa de disparitats serà simplement una imatge digital en escala de grisos tal que la intensitat de gris de cada píxel serà igual a la seva disparitat. Formalment:

Definició 9. *Sigui* (Ω, I) *una imatge digital en escala de grisos o en color. Donada una distància entre càmeres,* B*, el mapa de disparitats de* (Ω, I) *és la imatge digital en escala de grisos* (Ω, disp)

Figura 2.2: Mapa de disparitats

2.2 Obtenció del mapa de disparitats

Tot i que el concepte teòric de disparitat sigui molt senzill de definir, el seu càlcul efectiu pot ser molt complicat o no poder-se fer, ja que caldrà tenir dues imatges de la mateixa escena i aconseguir identificar els píxels que representin el mateix punt. Hi ha tècniques molt acurades però sovint no realitzables per falta de mitjans, com seria la obtenció de les profunditats de cada punt amb làser o amb llum estructurada. També hi ha algoritmes molt sofisticats, com el presentat a Sun [6]. Però en aquest projecte es proposarà un mètode molt senzill i que resultarà barat computacionalment parlant perqué s'implementarà d'una forma adequada.

2.2.1 Funcions de cost. Mètriques en imatges

Suposant que es tenen dos imatges d'una mateixa escena ja rectificades, es pren una de les imatges com a referència i s'intenta associar cada píxel de la imatge de referència amb un de l'altra imatge que estigui a la mateixa alçada (sobre la mateixa línia epipolar, que com les imatges estan rectificades és equivalent a dir que els píxels estan a la mateixa línia horitzontal). Mesurant el desplaçament que hi ha entre un píxel i el seu homòleg s'obté la disparitat de cada píxel. El criteri per a decidir quan dos píxels son homòlegs és la minimització d'una funció de cost que es construirà a partir d'una distància

entre imatges. Una manera natural de definir la distància entre dues imatges és mitjançant les normes L^p.

Definició 10. *Siguin* (Ω, I_1) *i* (Ω, I_2) *dues imatges digitals en escala de grisos definides sobre un mateix* Ω*. La distància en norma p és*

$$d_p((\Omega, I_1), (\Omega, I_2)) = \left(\sum_{(x,y)\in\Omega} |I_1(x,y) - I_2(x,y)|^p \right)^{1/p}$$

Es pot definir també la distància entre dos imatges amb conjunts de píxels diferents, (Ω_1, I_1) i (Ω_2, I_2), sempre i quan el conjunt de píxels d'una imatge sigui una traslació del conjunt de píxels de l'altra, es a dir si $\Omega_2 = (t_x, t_y) + \Omega_1$. Això significa que qualsevol element $(x_2, y_2) \in \Omega_2$ es pot escriure com $(x_2, y_2) = (t_x, t_y) + (x_1, y_1)$, amb $(x_1, y_1) \in \Omega_2$. La distància es defineix aleshores com

$$d_p((\Omega_1, I_1), (\Omega_2, I_2)) = \left(\sum_{(x,y)\in\Omega} |I_1(x,y) - I_2(t_x + x, t_y + y)|^p \right)^{1/p}$$

Aquestes definicions es poden extendre de manera natural a les imatges en color, de la següent manera:

$$d_p((\Omega_1, I_1), (\Omega_2, I_2)) = \left(\sum_{(x,y)\in\Omega} \|I_1(x,y) - I_2(t_x + x, t_y + y)\|_p^p \right)^{1/p}$$

on $\|\cdot\|_p$ és la norma p de $\mathbf{R}^3$, ja que I_1 i I_2 prenen valors en $\mathbf{R}^3$ (de fet en $\mathbf{Z}^3$) per tractar-se d'imatges en color.

Un cop s'ha donat una estructura mètrica a les imatges, ja es tenen les eines necessàries per a definir una funció de cost que permeti determinar quan dos píxels són homòlegs. Per a establir la relació d'homologia entre píxels no n'hi haurà prou amb exigir que siguin iguals o s'assemblin, s'haurà de demanar també que els seus entorns s'assemblin. Considerant com a referència la imatge dreta (Ω, IR), donat un píxel $(x, y) \in \Omega$, la estratègia inicial consistirà en prendre una finestra $W(x, y)$ al voltant del píxel, i escombrar tots els píxels (x', y) de la imatge esquerra (Ω, IL) que estiguin a la mateixa alçada y que el píxel original, considerant al voltant de (x', y) una finestra $W(x', y)$ que sigui la traslació de $W(x, y)$ al punt (x', y), és

a dir $W(x', y) = (x' - x, 0) + W(x, y)$. Els píxels homòlegs seran aquells que facin mínima la distància entre les restriccions de (Ω, IR) i (Ω, IL) a $W(x, y)$ i $W(x', y)$ respectivament, i la disparitat serà $\text{disp}(x, y) = x' - x$. Amb aquesta motivació es defineix la funció de cost:

$$F(x, x', y) = d_p((W(x, y), IR), (W(x', y), IL))$$

O, equivalentment,

$$F(x, x', y) = \left(\sum_{(\tilde{x},\tilde{y}) \in\ W(x,y)} \|IR(\tilde{x}, \tilde{y}) - IL(x' - x + \tilde{x}, \tilde{y})\|_p^p \right)^{1/p}$$

Així doncs, el problema es redueix a cercar, per cada píxel (x, y), el valor de x' que minimitza la funció anterior, i assignar el valor de disparitat $\text{disp}(x, y) = x' - x$. Pot passar, però, que dos píxels que teòricament haurien de ser homòlegs presentin valors molt diferents provocats per un canvi brusc en la il·luminació. És per això que abans de res és recomanable amitjanar els valors de IR i IL. O sigui que proporcionarà millors resultats minimitzar la funció

$$F(x, x', y) = \left(\sum_{(\tilde{x},\tilde{y}) \in\ W(x,y)} \|\bar{IR}(\tilde{x}, \tilde{y}) - \bar{IL}(x' - x + \tilde{x}, \tilde{y})\|_p^p \right)^{1/p}$$

on

$$\bar{IR} = \frac{IR}{\sum_{(\tilde{x},\tilde{y}) \in\ W(x,y)} IR(\tilde{x}, \tilde{y})}, \quad \bar{IL} = \frac{IL}{\sum_{(\tilde{x},\tilde{y}) \in\ W(x,y)} IL(\tilde{x}, \tilde{y})}$$

En definitiva, sigui quina sigui la funció que es minimitza, es tracta de resoldre un problema d'optimització combinatòria, per tant abans de res és convenient mirar si es pot reduir l'espai de solucions. En efecte, es poden tenir en compte les següents consideracions que simplifiquen bastant l'espai de recerca:

- La disparitat sempre té un valor més gran o igual que 0, per tant s'exploraran només aquells valors de x' tals que $x' \geq x$.
- Si dos píxels són molt diferents no poden ser homòlegs, per tant ja no s'exploraran les respectives finestres. Per decidir si dos píxels són molt diferents, s'imposa que la norma de la diferència dels seus valors sigui més gran que un cert llindar: $\|IR(x, y) - IR(x', y)\| > TOL$.

- Si a la escena no apareixen punts excessivament propers a les càmeres, es pot suposar que les disparitats no sobrepassen una certa disparitat màxima d_{max}, de manera que es restringirà la cerca a valors de x' tals que $x' \leq x + d_{max}$.

Un cop realitzat aquest pre-procès, la resolució del problema s'ha tornat molt més abordable. Aleshores, ja que es tracta d'un problema d'optimització combinatòria poden fer-se servir algunes tècniques habituals d'aquests problemes, com per exemple un algoritme de *branch and bound*. Però en aquest projecte s'ha optat per resoldre el problema mitjançant computació paral·lela, aprofitant la tecnologia CUDA incorporada en moltes tarjetes gràfiques de la casa nVidia.

2.2.2 Implementació en CUDA

Es pot aprofitar el fet que, degut a la naturalesa del problema, la seva resolució és independent per a cada píxel. Això vol dir que el càlcul de les disparitats de dos píxels diferents no es condicionen en absolut. Per tant, el temps de càlcul pot reduir-se molt si es disposa de la tecnologia apropiada per a calcular en paral·lel la disparitat per a cada píxel. Actualment no cal recórrer a súperordinadors, ja que des de la tardor de 2006 nVidia comercialitza tarjes gràfiques amb la tecnologia CUDA (Computer Unified Device Architecture), que permet aprofitar l'arquitectura del processador de la tarja gràfica per a realitzar càlculs en paral·lel. De fet, això es pot fer amb altres tarjes gràfiques sense la tecnologia CUDA, però el principal avantatge d'aquesta és la senzillesa del seu llenguatge de programació, una extensió del llenguatge C.

El procediment que se segueix és el següent:

- Copiar les dades de les imatges a la memòria de la tarja gràfica.
- Calcular en paral·lel la disparitat de cada píxel amb el processador de la tarja gràfica.
- Copiar el resultat a la memòria principal de l'ordinador.

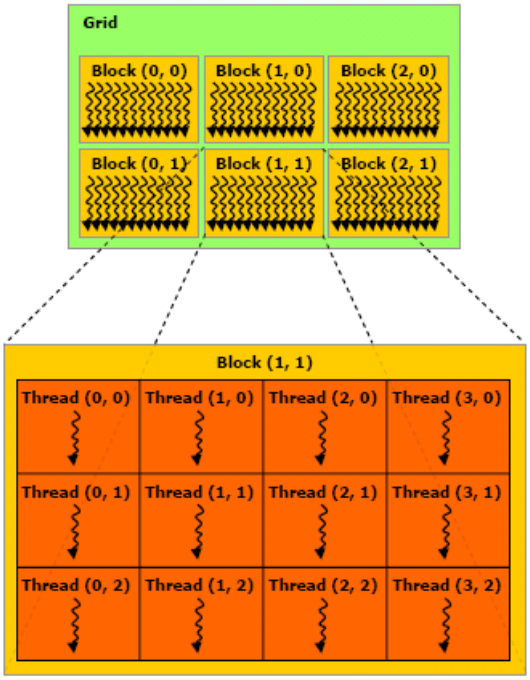


Figura 2.3: Grid, blocks i threads

Sense entrar en molt de detall sobre el funcionament de CUDA, la funció que s'executa paral·lelament, on es calcula la disparitat de cada píxel, s'anomena *kernel*. Els *kernels* s'executen als *threads*, que estan organitzats en *blocks* (matrius 1D, 2D o 3D de *threads*), que a la seva vegada estan organitzats en una *grid* (matriu 1D, 2D o 3D de *blocks*). L'estructura de *threads* i *blocks* pot triar-se lliurement dins d'unes certes limitacions; les dimensions màximes d'un *block* són $(512, 512, 64)$, però en total no poden haver-hi més de 512 *threads* per *block*. Cada block i cada thread té un identificador que permet repartir còmodament entre tots els threads la feina a executar.

En aquest projecte, els *threads* i els *blocks* s'han organitzat en matrius 2D, per tal de que quedin estructurats de la mateixa manera que una imatge, de manera que cada thread representi un píxel de manera natural. Per raons d'eficiència en l'aprofitament de *threads*, cada *block* s'ha estructurat com una matriu de 16×16 *threads*, i s'han usat el mínim de *blocks* per tal de cobrir tota la imatge. L'inconvenient d'aquesta estructura és que, generalment, s'estaran creant més threads dels necessaris, ja que les dimensions d'una imatge no necessàriament seran múltiples de 16. Per solucionar això hi ha dos alternatives. La primera és posar dins de cada *kernel* un

condicional, `if`, que comprovi si al *thread* en qüestió li correspon un píxel, i en cas que no li correspongui cap píxel no es realitzi el càlcul de la disparitat. La segona alternativa és augmentar la mida de la imatge afegint una files i/o columnes negres fins a aconseguir que les dimensions de la imatge quadrin amb el número de threads. La segona alternativa és la més eficient computacionalment parlant, ja que un condicional afegeix molta càrrega computacional a un *kernel*. L'incovenient és que el temps de transferència de dades entre la memòria de l'ordinador i la de la tarja gràfica fa que el procès no sigui tan ràpid com hauria de ser-ho teòricament, i això pot fer que no pugui realitzar-se el càlcul de mapes de disparitats a vídeos a temps real, tot i que lògicament per a aplicacions que no requereixin de ser realitzades a temps real és una molt bona solució. A Congote [5] es proposa un mètode per a calcular els mapes de disparitats mitjançant un algoritme de programació dinàmica fent servir CUDA.

Capítol 3

Generació d'imatges basada en mapes de disparitats

En aquest capítol s'explicarà la tècnica en la que es fonamentarà el desenvolupament del projecte. Donada una determinada imatge, la idea general consisteix en recuperar la profunditat a l'espai dels punts representats a cada píxel (això serà possible gràcies a la informació que proporcionarà el mapa de disparitats) i aplicar un moviment a aquests punts per a simular un moviment de la càmera i així aconseguir una segona imatge de la mateixa escena. Es veuran els problemes que comporta aquesta tècnica i es veuran diferents possibilitats per a solucionar-los.

3.1 Generació de noves imatges

Donada una imatge digital i el seu mapa de disparitats, la intenció és generar una nova imatge recuperant les coordenades espaials dels punts representats a cada píxel de la imatge, aplicar un moviment rígid a la posició de la càmera, i projectar novament els punts sobre el pla de la imatge. La referència serà el paper de Fehn [3]. Per comoditat en comptes de moure la càmera es mouran els punts de l'escena, ja que com en el model matemàtic de la càmera fotogràfica es requereix que aquesta estigui posicionada a l'ori-

gen de coordenades, si s'aplica el moviment a la càmera caldrà fer un canvi de coordenades per tal de tornar a situar l'origen sobre aquesta, en canvi si es mouen els punts de l'escena la càmera seguirà estant situada a l'origen. S'haurà de tenir en compte, això sí, que si el moviment desitjat per a la càmera és f, el moviment que caldrà aplicar als punts de l'escena per a aconseguir l'efecte equivalent haurà de ser f^{-1}.

3.1.1 Traslacions

Les traslacions són el moviment més simple de tots. De fet, per a generar imatges estereoscòpiques es partirà d'una imatge digital (Ω, I) que es farà servir per a un ull i mitjançant una traslació horitzontal es generarà una nova imtage (Ω, J) que es farà servir per a l'altre ull.

Com ja s'ha comentat anteriorment, si el que es desitja és aplicar a la càmera una traslació $f(X,Y,Z) = (X,Y,Z)+T$ el que es farà realment serà aplicar el moviment f^{-1} als punts de l'escena. En el cas d'una traslació, el moviment invers és especialment senzill: $f^{-1}(X,Y,Z) = (X,Y,Z) - T$.

La imatge traslladada s'obtindrà aplicant la transformació f^{-1} punt a punt. Així, doncs, el procediment per a obtenir la imatge traslladada és el següent:

- Recuperació de les coordenades espaials del punt representat al píxel (x,y):

$$(X,Y,Z) = \left(x\frac{fB}{\text{disp}(x,y)}, y\frac{fB}{\text{disp}(x,y)}, \frac{fB}{\text{disp}(x,y)}\right)$$

- Càlcul de les coordenades del punt traslladat:

$$(\bar{X},\bar{Y},\bar{Z}) = (X,Y,Z) - T$$

- Projecció del punt traslladat sobre el pla de la imatge:

$$(\bar{x},\bar{y}) = \left(f\frac{\bar{X}}{\bar{Z}}, f\frac{\bar{Y}}{\bar{Z}}\right)$$

  En el cas en que sobre un mateix píxel (x,y) es projecti més d'un punt, s'agafarà aquell que estigui a menor profunditat.

- La nova imatge serà (Ω, J), on $J(\bar{x}, \bar{y}) = I(x, y)$

Si la traslació que s'aplica és únicament horitzontal, per a generar una imatge estereoscòpia, les expressions anteriors esdevenen molt simples. En particular, les noves coordenades del píxel (x, y) són

$$(\bar{x}, \bar{y}) = \left(x - \frac{h}{B}\text{disp}(x, y), y \right)$$

on h és la traslació horitzontal.

De tot aquest procediment es dedueix que és imprescindible el coneixement de la distància entre càmeres, B, a partir del qual s'ha obtingut el mapa de dispariats. A mode d'avançament, es comenta que la forma en que s'aplicarà aquesta tècnica a la laparoscòpia consistirà en introduir dues càmeres separades una distància B prou petita com per a que el dispositiu no sigui excessivament voluminós, però que serà insuficient per a poder obtenir una imatge estereoscòpica adequada. No obstant, es podrà calcular el mapa de disparitats, i a partir d'aquí es podrà generar una nova imatge que serà la que permetrà obtenir una visió estereoscòpia correcta.

El principal inconvenient d'aquest procediment és que hi haurà píxels de la nova imatge sobre els quals no es projectarà cap punt. Aquests píxels es coneixen com a forats o disoclusions. El problema de com omplir amb informació aquests forats es tractarà més endavant. De moment, si no existeix cap (x, y) que al fer el moviment es converteixi en $(\bar{x}, \bar{y})$, aleshores $J(\bar{x}, \bar{y}) = 0$ si la imatge és en escala de grisos i $J(\bar{x}, \bar{y}) = (0, 0, 0)$ si la imatge és en color, o el que és el mateix, es pinten els forats amb color negre.

3.1.2 Rotacions

En alguns casos pot ser interessant aplicar una rotació a la imatge, per exemple per rectificar-la si és necessari. Novament, si el que es desitja és aplicar una rotació f a la càmera, el que es farà serà aplicar la rotació f^{-1} als punts de l'escena. Obtenir la inversa d'una rotació també és molt senzill, ja que segons s'ha vist al primer capítol, la matriu R d'una isometria lineal verifica que $R^{-1} = R^t$. També es pot obtenir directament la matriu de f^{-1} tenint en compte que la inversa d'una rotació de α radians al voltant

d'un eix amb vector director v és una rotació de $-\alpha$ radians al voltant del mateix eix. El procediment a seguir és idèntic al que s'ha seguit per a les traslacions:

- Recuperació de les coordenades espaials del punt representat al píxel (x, y):
$$(X, Y, Z) = \left(x\frac{fB}{\text{disp}(x,y)}, y\frac{fB}{\text{disp}(x,y)}, \frac{fB}{\text{disp}(x,y)} \right)$$
- Càlcul de les coordenades del punt rotat:
$$(\bar{X}, \bar{Y}, \bar{Z}) = f^{-1}(X, Y, Z)$$
- Projecció del punt rotat sobre el pla de la imatge:
$$(\bar{x}, \bar{y}) = \left(f\frac{\bar{X}}{\bar{Z}}, f\frac{\bar{Y}}{\bar{Z}} \right)$$
En el cas en que sobre un mateix píxel (x, y) es projecti més d'un punt, s'agafarà aquell que estigui a menor profunditat.
- La nova imatge serà (Ω, J), on $J(\bar{x}, \bar{y}) = I(x, y)$

Igual que amb les traslacions, es generaran forats que caldrà omplir.

3.1.3 Moviment general

El moviment més general que s'aplicarà a una imatge serà una rotació no lineal, però com ja s'ha vist, tot moviment rígid afí es pot descomposar com a una isometria lineal seguida d'una traslació. Per tant, si f és una rotació no lineal es té la següent descomposició:
$$f(X, Y, Z) = R(X, Y, Z)^t + T^t$$
on R és la matriu de la rotació lineal en la base ordinària.

Per tant, la transformació inversa, que és la que s'aplicarà als punts de l'escena, és
$$f^{-1}(X, Y, Z) = R^t \left[(X, Y, Z)^t - T^t \right]$$
Un cop moguts els punts, el procediment a seguir és idèntic al que se segueix per les traslacions i per les rotacions. Novament, es crearan forats en la imatge que s'hauran d'omplir.

3.2 Ompliment de forats

Un cop vist com generar una nova imatge a partir de la imatge original i el seu mapa de disparitats, es discutiran diferents alternatives per a l'ompliment dels forats que es generen en la imatge. Es presentaran totes les tècniques que s'han assajat en aquest projecte, tant les que s'han fet servir per a l'aplicació final com les que no. Per a decidir quines tècniques es fan servir, interessarà trobar un equilibri entre qualitat dels resultats i velocitat de càlcul.

3.2.1 Definició del concepte de forat

Tot i que intuitivament ja s'ha presentat el concepte de forat, abans de continuar cal formalitzar-lo per a poder-ne parlar rigurosament. Com s'ha dit, la mesura adoptada fins al moment amb els forats ha estat deixar-los de color negre. Per tant, sembla lògic definir els forats com aquells píxels que són negres, però també es pot donar el cas que un píxel sigui negre perqué provingui d'un píxel de la imatge original que també era negre, de manera que els píxels negres d'aquest tipus cal excloure'ls de la definició de forat.

Definició 11. *Sigui* (Ω, I) *una imatge digital en color o en escala de grisos de la qual es coneix un mapa de disparitats, i* (Ω, J) *la imatge digital obtinguda al aplicar un moviment rígid* $f : \boldsymbol{R}^3 \to \boldsymbol{R}^3$*. Si* (x, y) *és tal que* $I(x, y) = 0$ *o* $(0, 0, 0)$*, el píxel* $(\bar{x}, \bar{y})$ *corresponent a* (x, y) *després d'aplicar el moviment rep el nom de fals forat. Un píxel* $(\bar{x}, \bar{y})$ *es dirà que és un forat si* $J(\bar{x}, \bar{y}) = 0$ *o* $(0, 0, 0)$ *i no és un fals forat.*

De manera que el procediment per detectar forats és simple:

- Es busquen tots els píxels negres en la imatge original i el seu píxel corresponent a la nova imatge s'etiqueta com a fals forat.
- Es busquen tots els píxels negres en la imatge nova i aquells que no estiguin etiquetats com a fals forat s'etiqueten com a forats.

3.2.2 Mida d'un forat

Un cop s'ha aconseguit detectar els forats, el següent pas consisteix en mesurar la seva mida per tal de poder utilitzar diferents estratègies per a omplir els forats depenent de si són més grans o més petits. A mode d'avançament, tot i que és fàcilment imaginable, els forats s'ompliran usant la informació continguda en els píxels més propers que no siguin forats (interpolació i inpainting). Per tant, és fàcil imaginar-se que l'efecte aconseguit quan s'omple un forat serà millor quant menys distància hi hagi als píxels amb informació. Així doncs, donat un forat $(\bar{x}, \bar{y})$ es tracta de mesurar la distància L_h entre els dos píxels no forats més propers per l'esquerra i per la dreta de $(\bar{x}, \bar{y})$, i la distància L_v entre els dos píxels més propers per sobre i per sota.

Definició 12. *Sigui $(\bar{x}, \bar{y})$ un forat i $L_h(\bar{x}, \bar{y})$ i $L_v(\bar{x}, \bar{y})$ les distàncies definides al paràgraf anterior. Es defineix la mida del forat $(\bar{x}, \bar{y})$ com*

$$\mu(\bar{x}, \bar{y}) = min\{L_h(\bar{x}, \bar{y}), L_v(\bar{x}, \bar{y})\}$$

Aquesta mesura de la grandària d'un forat es pot extendre facilment a qualsevol subconjunt de Ω que estigui format per forats. Si $H \subset \Omega$ és un conjunt de forats, la seva mida es defineix com

$$\mu(H) = \sum_{(\bar{x},\bar{y}) \in H} \mu(\bar{x}, \bar{y})$$

de manera que té sentit parlar també de mides de conjunts de forats.*

3.2.3 Interpolació

La interpolació lineal és el mètode més senzill per a omplir forats. A la figura 3.1 s'il·lustra el procediment en el cas d'una imatge digital en escala de grisos.

En el cas d'una imatge digital en color, el procediment consisteix en interpolar cada canal de color per separat (3.2).

*Si H_Ω és el conjunt format per tots els forats que hi ha a Ω, aleshores $(H_\Omega, \mathcal{P}(H_\Omega), \mu)$ és un espai de mesura

Figura 3.1: Interpolació de forats en escala de grisos

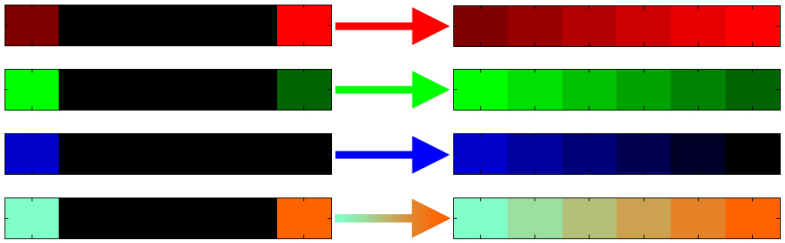


Figura 3.2: Interpolacio d'una imatge en color

Figura 3.3: Interpolació de forats

Donat un forat $(\bar{x}, \bar{y}) \in \Omega$, la interpolació es farà sempre en la direcció que determini la mida del forat. És a dir, si $\mu(\bar{x}, \bar{y}) = L_h(\bar{x}, \bar{y})$ s'interpolarà en direcció horitzontal, i si $\mu(\bar{x}, \bar{y}) = L_v(\bar{x}, \bar{y})$ s'interpolarà en direcció vertical.

El major inconvenient de la interpolació és que es pot donar el cas de que s'estigui interpolant amb píxels que no pertanyin a un mateix objecte,

Figura 3.4: Mescla d'objectes en la interpolació

de manera que els forats s'estan omplint amb una mescla d'informació dels dos objectes. A la figura 3.4 es mostra aquest efecte.

3.2.4 Inpainting

Una altra tècnica per a omplir forats és l'inpainting. Aquesta tècnica és la que fan servir els artistes i restauradors per a restaurar obres d'art deteriorades. Es coneix com a inpainting digital la modelització matemàtica i posterior implementació algorísmica d'aquestes tècniques. A [10] es presenta amb detall el procediment. A la figura 3.5 s'han esborrat les lletres mitjançant inpainting.

La implementació del mètode descrit a [10] proporciona bons resultats però és massa costosa computacionalment. A [11] es proposa un algoritme d'inpainting ràpid que, tot i proporcionar resultats menys acurats, és molt més senzill i eficient, de manera que la reducció del temps de càlcul compensa la pèrdua de qualitat.

L'inpainting ràpid consisteix simplement en fer la convolució de la imat-

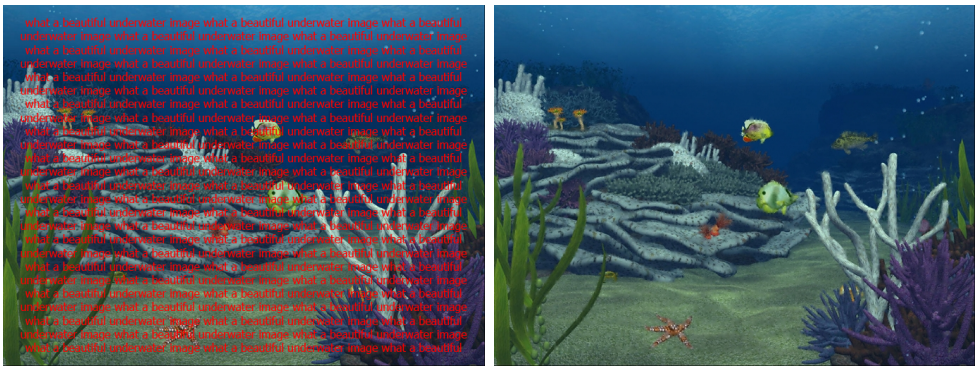


Figura 3.5: Inpainting

ge per un dels següents nuclis, reiteradament fins arribar al nivell de convergència desitjat o fins assolir el nombre màxim d'iteracions:

$$\begin{bmatrix} a & b & a \\ b & 0 & b \\ a & b & a \end{bmatrix} \begin{bmatrix} c & c & c \\ c & 0 & c \\ c & c & c \end{bmatrix}$$

amb $a = 0.073235$, $b = 0.176765$, $c = 0.125$. Usant el primer d'aquests dos nuclis, s'ha pogut comprovar que tot i que l'inpainting proporciona resultats una mica més agradables a la vista que la interpolació, no se soluciona el problema de la mescla d'objectes i a més el temps de càlcul és més gran que amb la interpolació, de manera que tampoc compensa.

Figura 3.6: Inpainting ràpid

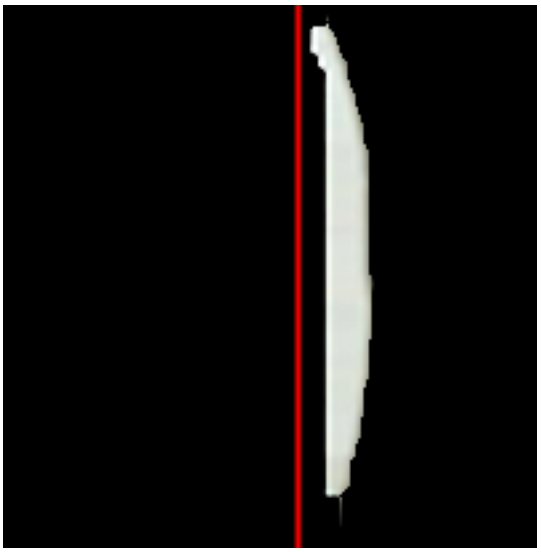


Figura 3.7: Imatge amb informació insuficient per interpolar

3.2.5 Interpolació per capes

Un primer intent per a solucionar el problema de la mescla d'objectes consisteix en interpolar a cada capa de profunditat per separat, interpolant per capes s'evitarà mesclar-los.

El problema és que el nombre de capes de profunditat pot ser molt gran, per exemple en una imatge poden haver-hi des de objectes que estiguin a pocs centímetres de profunditat (herbes, flors, branques,...) a d'altres que estiguin a quilòmetres de profunditat (per exemple, muntanyes), per tant, caldrà decidir quin gruix tindrà cada capa de profunditat, i si s'opta per donar poc gruix sortiran poques capes i alguns objectes se seguiran mesclant, en canvi si es decideix fer les capes més fines aleshores el temps de càlcul pot créixer molt ràpidament. Un altre inconvenient que pot sorgir és que un forat quedi a la frontera d'una capa de profunditat, i per tant no hi hagi suficient informació per interpolar, com es pot veure a la figura 3.7, on la línia vermella marca la frontera de la capa de profunditat on es troba l'objecte blanc, de manera que al interpolar el forat existent entre la frontera i l'objecte, no hi haurà prou informació.

A la figura 3.8 es pot veure el resultat fent servir 500 capes de profunditat.

Figura 3.8: Interpolació amb 500 capes de profunditat

Com es pot veure en els forats que s'interpolen el resultat és força bo, però a part de que el temps de càlcul no compensa, queden forats sense omplir justament degut a l'inconvenient exposat abans de que els forats puguin quedar a la frontera de la capa de profunditat.

3.2.6 Correcció del mapa de disparitats

La següent estratègia consisteix en aplicar diverses correccions al mapa de disparitats. Per una banda es veurà el filtrat gaussià, a partir del qual el que s'intenta és disminuir la mida dels forats, i per altra banda es veurà la dilatació, que té com a finalitat aconseguir que els forats no es generin entre dos objectes, per evitar que a l'hora d'interpolar es mesclin objectes diferents.

Filtrat gaussià

El filtrat gaussià del mapa de disparitats és el mètode proposat a Zhang [7], Tam [8] i Lee [13]. Consisteix en fer un filtrat del mapa de disparitats mitjançant una convolució de nucli

$$g(x, y, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2 + y^2}{\sigma^2}\right)$$

amb $-w \leq x \leq w$ i $-w \leq y \leq w$ per a un cert $w \in \mathbf{Z}^+$. A [7] es proposa que $w = 1.5\sigma$ ja que empíricament s'ha comprovat que aquest valor funciona

bé, i es proposa també que $\sigma = 30$.

Provant aquest mètode amb el primer nucli, s'ha comprovat que, efectivament, es redueixen les mides dels forats tal com s'afirma a [7], però al difuminar el mapa de disparitats s'estan alterant les profunditats reals dels punts representats, de manera que a la nova imatge generada els objectes poden sortir distorsionats, o bé pot passar que els objectes no es moguin fins a la posició real que haurien d'ocupar si no s'haguessin alterat les profunditats. Per al primer dels problemes, la solució que es proposa a [7] és fer servir el següent nucli asimètric:

$$g(x, y, \sigma_h, \sigma_v) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{\sigma_h^2} - \frac{y^2}{\sigma_v^2}\right)$$

triant σ_h i σ_v adequadament per evitar la distorsió. De totes formes, no es proposa ni es suggereix cap mètode per a escollir els valors de σ_h i σ_v adequats per a una imatge en general qualsevol. Per tant és molt probable que, depenent de la geometria dels objectes representats a la imatge, s'hagin de triar uns valors molt diferents.

El problema més greu és l'alteració de les posicions reals dels objectes degut a la pèrdua de resolució de profunditats, tal com es pot apreciar a la figura 3.9, on la imatge superior representa la traslació sense haver tractat el mapa de disparitats, en la qual es pot apreciar la posició real fins a la que hauria d'arribar la bitlla, i la imatge inferior representa la traslació després d'haver aplicat un filtre gaussià al mapa de disparitats, on es veu com s'han reduit notablement les mides dels forats però s'ha alterat la posició dels objectes, tal com es pot comprovar veient que la bitlla ja no ocupa la posició que hauria d'ocupar.

Dilatació i interpolació del mapa de disparitats

Per tal d'evitar la mescla d'objectes, la solució que es proposa en aquest projecte consisteix en fer una dilatació del mapa de disparitats. Amb això el que s'aconsegueix és que quan hi ha dos objectes adjacents a diferents profunditats (equivalentment, que tenen diferents disparitats), s'enganxen uns píxels de l'objecte llunyà (el de menys disparitat) a l'objecte proper (el de més disparitat) de manera que al fer la traslació, un petit fragment de l'objecte llunyà romandrà enganxat a l'objecte proper i així el forat es ge-

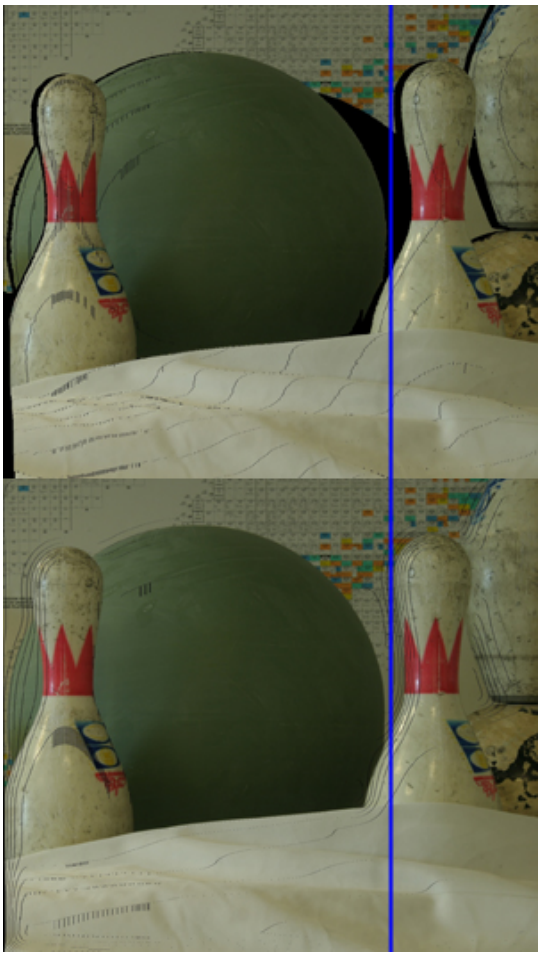


Figura 3.9: Alteració de les posicions reals

nerarà únicament entre píxels de l'objecte llunyà, evitant que a l'interpolar es produeixi una barreja d'objectes. A més a més, per tal d'evitar discontinuitats massa brusques produides pels punts de disparitat desconeguda (aquells que tenen valor de disparitat 0), s'ompliran aquests punts interpolant amb els valors de les disparitats dels píxels veins. En la figura 3.10 es pot apreciar l'efecte de dilatar i interpolar el mapa de disparitats i el resultat que produeix en la nova imatge generada i en la posterior interpolació d'aquesta.

Com es pot comprovar, un tros de l'objecte blanc queda enganxat a l'objecte marró que està a menys profunditat, de manera que el forat es produeix únicament entre píxels de l'objecte blanc i per tant ja no es produeix la mescla d'objectes que estiguin a diferents capes de profunditat.

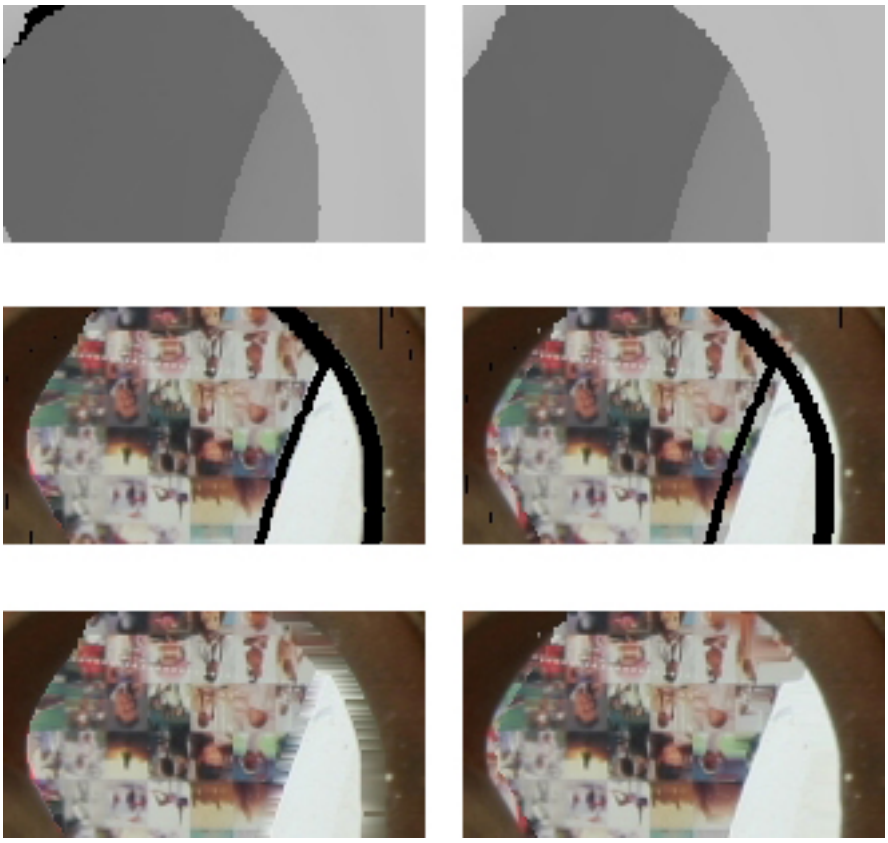


Figura 3.10: Dilatació del mapa de disparitats

També succeeix el mateix amb l'objecte blanc i el dibuix que hi ha darrere seu.

Aquesta tècnica preesnta principalment dos inconvenients. Un d'ells, més que un inconvenient és una restricció, és que cal fer la traslació en la mateixa direcció i sentit que s'ha usat per generar el mapa de disparitats. En cas de no fer-se així, pot passar que alguns dels píxels que s'han enganxat a un objecte de menor profunditat trepitjin aquest objecte al fer la traslació, com succeeix en la figura 3.11. Com s'ha dit, això no és un inconvenient si se sap en quin sentit s'ha mogut la càmera per a generar el mapa de disparitats, tant sols és una restricció.

L'altre inconvenient, el major dels dos, és el cas que s'il·lustra en la figura 3.12.

Els objectes 1 i 2 estan a la mateixa profunditat (tenen la mateixa

Figura 3.11: Sentit incorrecte de la traslació

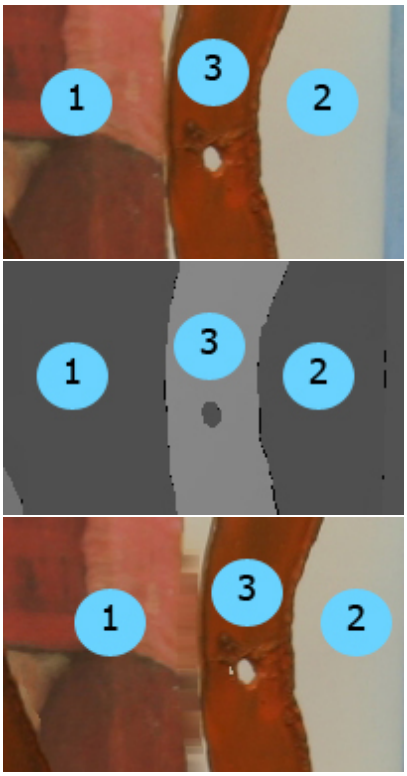


Figura 3.12: Defectes de la dilatació del mapa de disparitats

disparitat), en canvi l'objecte 3 està per davant. Al dilatar la capa de profunditat de l'objecte 3, s'està enganxant a aquest un petit fragment de l'objecte 1, de manera que al generar la nova imatge es crea un forat entre l'objecte 1 i l'objecte 2 i al fer la interpolació es mesclen encara que estiguin a la mateixa profunditat. Els punts a favor són la senzillesa i rapidesa de la implementació del mètode, i el fet que no distorsioni ni disminueixi la resolució de la profunditat.

3.2.7 Comparació entre les diferents alternatives

En la següent taula es compararan els diferents mètodes assajats per poder veure els avantatges i inconvenients dels uns sobre els altres:

	Qualitat	**Velocitat**	**Senzillesa**
Interpolació	Dolenta	Bona	Bona
Inpainting	Regular	Molt dolenta	Molt bona
Interpolació per capes	Regular	Dolenta	Dolenta
Filtrat gaussià + Interpolació	Regular	Bona	Bona
Dilatació + Interpolació	Bona	Bona	Bona

Cal remarcar que l'avaluació d'aquests s'ha fet a partir de les implementacions d'aquests mètodes que ha fet l'autor del projecte. És posible que hi hagi altres implementacions que millorin o empitjorin alguns dels aspectes tractats. El que sembla bastant evident, però, és que no hi ha cap mètode que proporcioni una qualitat perfecta, tots tenen algun defecte, però el que queda patent és que el fet de tractar el mapa de disparitats proporciona resultats favorables. El mètode escollit serà el consistent en dilatar i interpolar el mapa de profunditats i omplir els forats per interpolació.

3.3 Resum del procediment utilitzat

S'acabarà aquest capítol combinant i ordenant les tècniques exposades anteriorment que es faran servir, per esquematitzar el procediment que s'utilitza per a generar les imatges estereoscòpiques.

- Es parteix d'una imatge original i el seu mapa de disparitats. Es pren la imatge original com la imatge que es farà servir per un dels ulls. La imatge de l'altre ull serà la que es generarà artificialment.

- Es dilata el mapa de disparitats i s'interpolen aquells píxels de disparitat 0.

- Es recuperen les profunditats dels punts representats a cada píxel a partir del mapa de disparitats.
- Es genera la nova imatge per traslació. Apareixen forats.
- S'omplen els forats per interpolació lineal.
- Es genera la imatge estereoscòpica combinant les imatges dels dos ulls.

Capítol 4

Composició

4.1 Tècniques de composició

La composició d'imatges estereoscòpiques és el procés de muntatge de les imatges esquerra i dreta per tal de ser visualitzades en tres dimensions. Es presentaran tres de les tècniques més habituals i es compararan entre elles. La idea comú a totes les tècniques és fer que l'ull esquerra únicament pugui veure la imatge esquerra i l'ull dret només la dreta.

4.1.1 Anaglif

L'anaglif és la tècnica més senzilla i barata, ja que no cal cap dispositiu especial ni per composar ni mostrar les imatges, i les ulleres que es necessiten no són gens sofisticades.

La idea consisteix simplement en filtrar les dues imatges amb colors oposats, generalment vermell i cyan, i en menor mesura verd i magenta. Les ulleres es construiran amb unes lents tals que cadascuna d'elles deixi veure tant sols el color amb que ha estat filtrada la seva imatge corresponent. Habitualment, es filtra la imatge esquerra amb color vermell i la dreta amb

Figura 4.1: Cercle de color

Figura 4.2: Ulleres anaglifes

cyan, de manera que la lent esquerra serà vermella i la dreta cyan, així l'ull esquerra no podrà veure la imatge dreta i viceversa.

El principal avantatge del sistema anaglif és que no calen dispositius ni ulleres sofisticades. Es poden crear anaglifs amb un simple full de paper i una impresora en color, o una pantalla d'ordinador, o un televisor, etc, i per visualitzar la imatge es poden construir unes ulleres fent servir tant sols paper de celofan vermell i cyan. Per tant, la composició anaglifa és una solució realment barata i per tant molt usada.

El major inconvenient és que al manipular els colors de les imatges s'està distorsionant informació. Així, per exemple, els objectes de color vermell no es veuran bé perque a la imatge dreta desapareixeran. En canvi, per a imatges en escala de grisos aquesta tècnica presentarà resultats pràcticament perfectes, ja que en aquest cas filtrar el vermell o el cyan en una imatge en escala de grisos no farà perdre informació. No obstant, tot i aquest inconvenient, el seu baix cost fa que sigui una tècnica àmpliament utilitzada per distribuir material visual estereoscòpic.

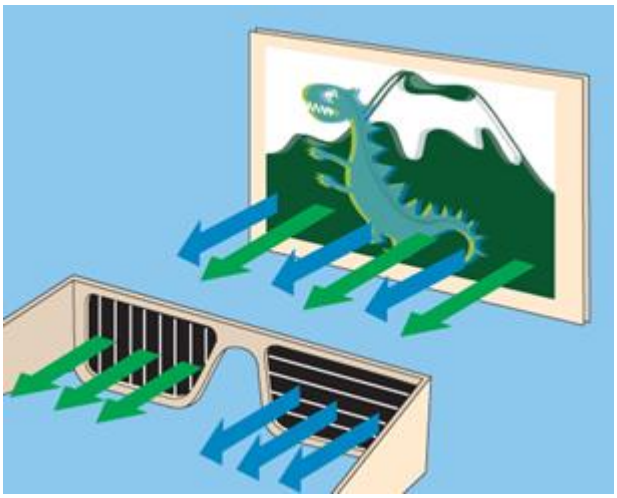


Figura 4.3: Polarització

4.1.2 Polarització

Es fa servir llum polaritzada per separar les imatges esquerra i dreta. En general es polaritza una imatge en plans horitzontals i l'altra en plans verticals. Les ulleres que es requereixen tenen a cada lent un filtre polaritzador adequat de forma que cada ull rebi la imatge correcta. És una solució relativament barata i els colors no es veuen afectats, encara que sí que es produeix una lleugera pèrdua de lluminositat.

Un petit inconvenient és que si l'usuari no manté els ulls amb una orientació correcta es perdrà la sensació de 3D. Això es pot solucionar fent servir polarització circular, per tal que la separació de les imatges sigui independent de la orientació. El major inconvenient és que si la imatge en comptes de ser emesa és projectada sobre una pantalla, caldrà que aquesta estigui construida amb algun material que mantingui la polarització de la llum. En cas de projectar-se sobre una pantalla convencional es perdrà la polarització i per tant l'efecte 3D.

4.1.3 Ulleres actives

Aquesta tècnica consisteix en emetre de forma alternada les imatges esquerra i dreta. Per visualitzar correctament la composició, cal fer servir unes

Figura 4.4: Ulleres polaritzades

ulleres de lents de cristall líquid (LCD) que obrin i tanquin la visió de cada ull de manera sincronitzada amb la pantalla emisora, és a dir, que mentre s'estigui emetent la imatge esquerra la visió de l'ull dret estigui tancada i viceversa. A altes freqüències, l'obertura i el tancament de la visió de cadascun dels ulls és imperceptible. Una freqüència de refresc habitual en moltes pantalles i monitors que emeten imatges normals és de $60Hz$, per tant moltes pantalles per emetre imatges 3D ho fan a $120Hz$, ja que han d'emetre el doble d'imatges per segon.

Figura 4.5: Ulleres actives

El principal avantatge és que no hi ha cap tipus de pèrdua de qualitat, ni de color ni de lluminositat. El major inconvenient és l'elevat preu, tant del dispositiu emisor com de la pantalla, a part de l'elevat consum de bateries que generen aquest tipus d'ulleres.

4.1.4 Comparació de les diferents tècniques

A continuació es presenta una taula comparant les diferents tècniques exposades:

	Qualitat	**Preu**
Anaglif	Acceptable	Molt econòmic
Polarització	Bona	Econòmic
Ulleres actives	Òptima	Car

Per senzillesa i economia en aquest projecte es farà servir la composició amb anaglifs, ja que no es disposen de mitjans per fer-ho amb polarització i la tècnica de les ulleres actives és massa cara.

4.1.5 Altres tècniques

Hi ha altres tècniques menys habituals, que es comentaran molt breument.

Una d'elles és la tècnica Cromatek, consistent aprofitar el fet que l'angle de refracció de la llum a través d'un prisma depèn de la longitud d'ona d'aquesta, és a dir del color. Per tant, el que es fa és codificar la profunditat amb colors. Les ulleres que es fan servir estan construides amb uns microprismes que refracten la llum. Al igual que l'anaglif, pot realitzar-se amb mitjans senzills, tot i que les ulleres són més costoses. Com a avantatge sobre l'anaglif té que la imatge pot visualitzar-se també en 2D sense les ulleres (en anaglif es veuen dues imatges superposades). Com a desventatge té que es perd totalment la informació cromàtica.

Una tècnica molt popular és la coneguda com a Head-mounted display (HMD). Consisteix en un casc que mostra les dos imatges per separat per cada ull.

Una altra tècnica l'autoestereoscopia. Amb aquesta tècnica no calen ulleres. Consisteix en una pantalla amb petites lents al damunt. De moment, tot i que s'han comercialitzat televisors amb aquesta tecnologia, és una tècnica encara en desenvolupament.

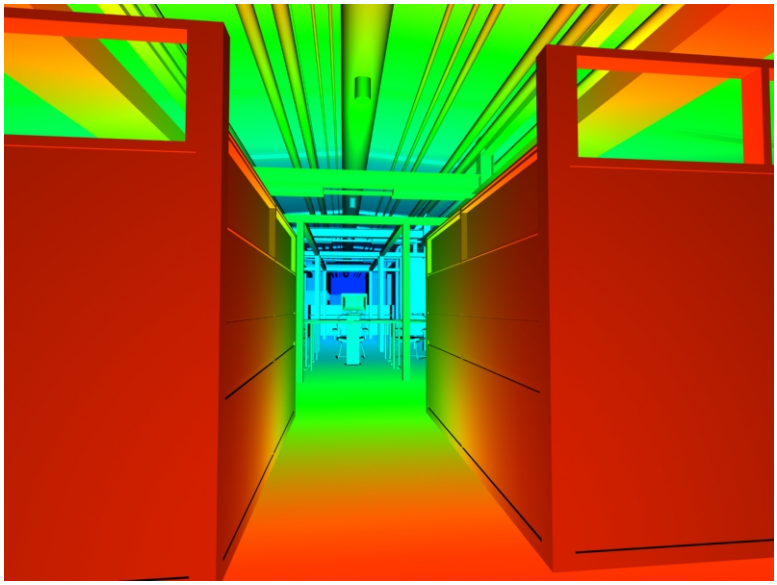


Figura 4.6: Imatge Cromatek

Capítol 5

Resultats

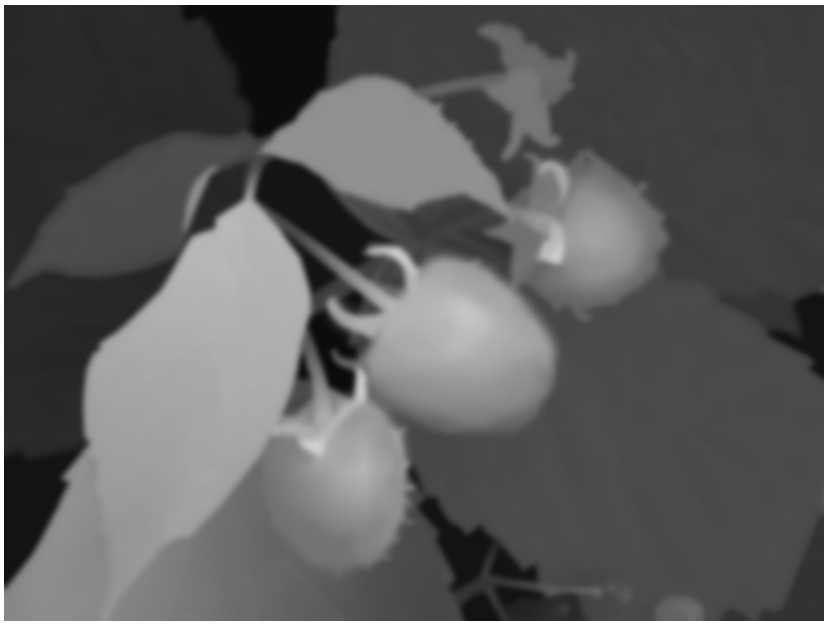
En aquest capítol es presentaran alguns resultats. Els càlculs s'han realitzat amb Matlab 7, en un procesador Intel®Pentium®processor T4500. Els mapes de disparitats s'han calculat amb una tarja gràfica NVIDIA GeForce 9500GT.

5.1 Mores

En aquesta imatge el mapa de disparitats s'ha agafat ja calculat. Es parteix de la imatge dreta i del mapa de disparitats i es genera la imatge esquerra.

5.1.1 Imatges

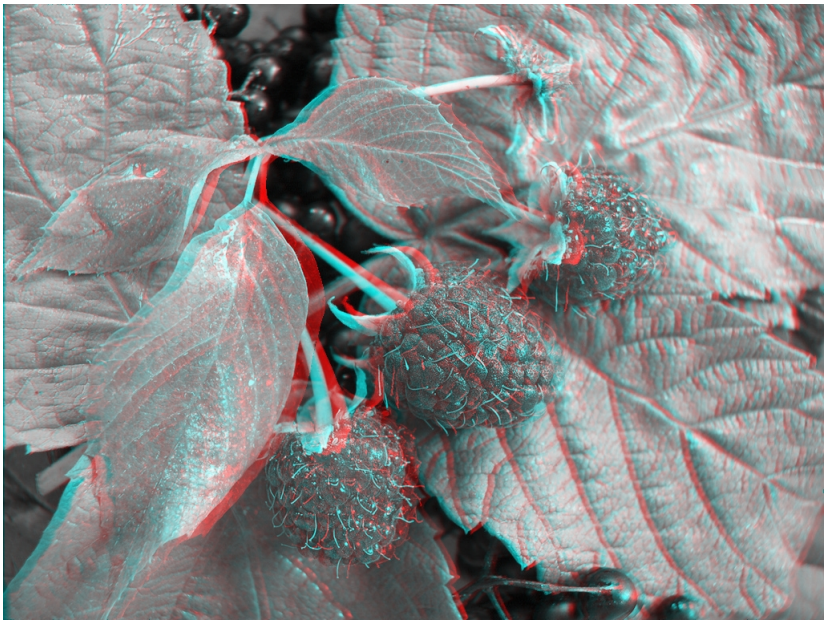
Imatge dreta original
Font: http://www.triaxes.com



Mapa de disparitats proporcionat per l'autor de la imatge
Font: http://www.triaxes.com

Imatge esquerra amb forats

Imatge esquerra amb forats tapats



Composició

Calen ulleres anaglifes per poder veure correctament la imatge

5.1.2 Temps de càlcul

Resolució de la imatge: 768×1024

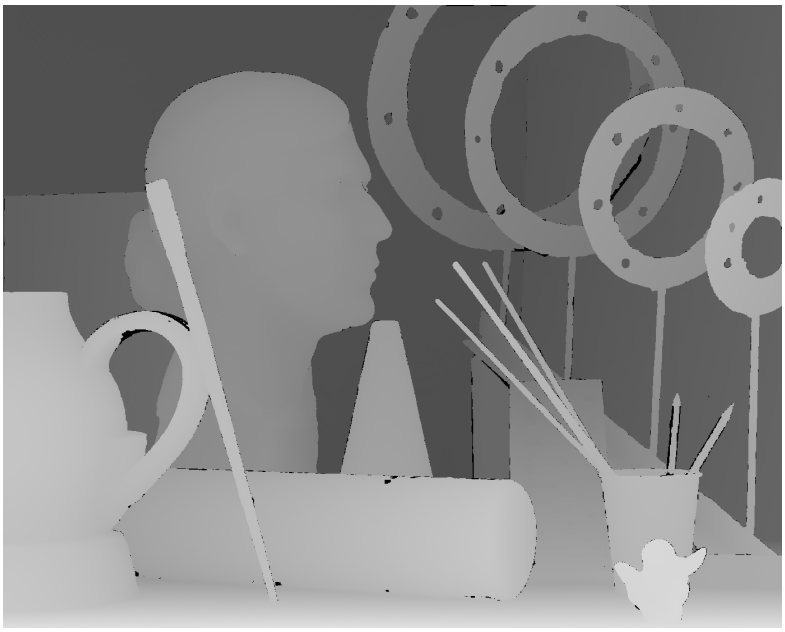
Tractament del mapa de disparitats	8.47s
Generació de la nova imatge	7.74s
Interpolació de forats	8.63s

5.2 Art

En aquest cas, novament s'ha agafat un mapa de disparitats ja proporcionat per l'autor de les imatges.

5.2.1 Imatges

Imatge dreta original
Font: http://http://vision.middlebury.edu/stereo/

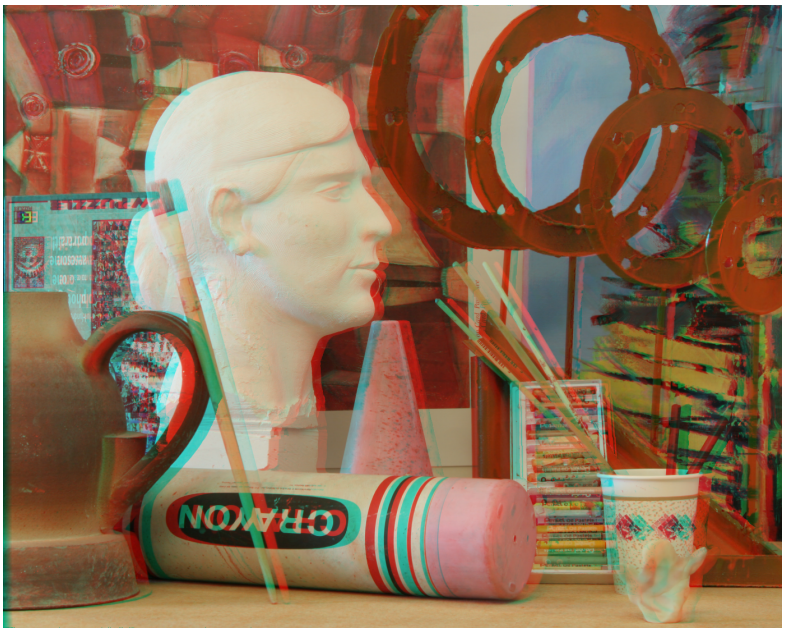


Mapa de disparitats proporcionat per l'autor de la imatge
Font: http://http://vision.middlebury.edu/stereo/

Mapa de disparitats corregit i tractat

Nova imatge esquerra amb forats

Nova imatge esquerra amb forats tapats



Composició Calen ulleres anaglifes per poder veure correctament la imatge

5.2.2 Temps de càlcul

Resolució de la imatge: 1110×1390

Tractament del mapa de disparitats	16.41s
Generació de la nova imatge	15.20s
Interpolació de forats	16.02s

5.3 Oset

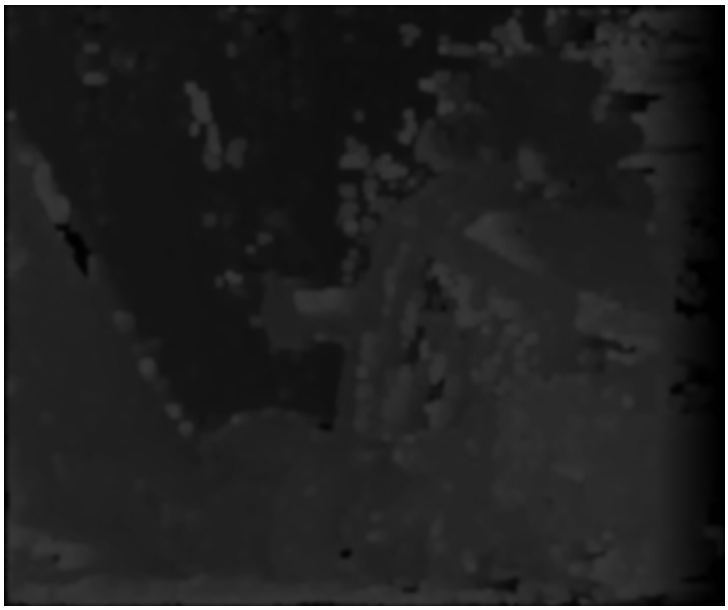
Per aquesta imatge es prenen 2 imatges de la mateixa escena, preses amb la orientació i alçada correcta de les càmeres (és a dir que no cal rectificar-les), però situades a una distància inapropiada per a generar una imatge estereoscòpica. Es calcularà el mapa de disparitats i es generarà una nova imatge esquerra de tal manera que la composició de les imatges doni lloc a una imatge estereoscòpica que doni sensació de 3 dimensions.

5.3.1 Imatges

Imatge esquerra original
Font: http://http://vision.middlebury.edu/stereo/

Imatge dreta original
Font: http://http://vision.middlebury.edu/stereo/

Mapa de disparitats calculat



Mapa de disparitats tractat i corregit

Nova imatge esquerra amb els forats tapats

Composició

Calen ulleres anaglifes per poder veure correctament la imatge

5.3.2 Temps de càlcul

Resolució de la imatge: 375×450

Càlcul del mapa de disparitats	20s (aprox.)
Tractament del mapa de disparitats	2.72s
Generació de la nova imatge	1.58s
Interpolació de forats	1.95s

Nota: El temps de càlcul de la tarja gràfica no s'ha pogut mesurar de manera precisa.

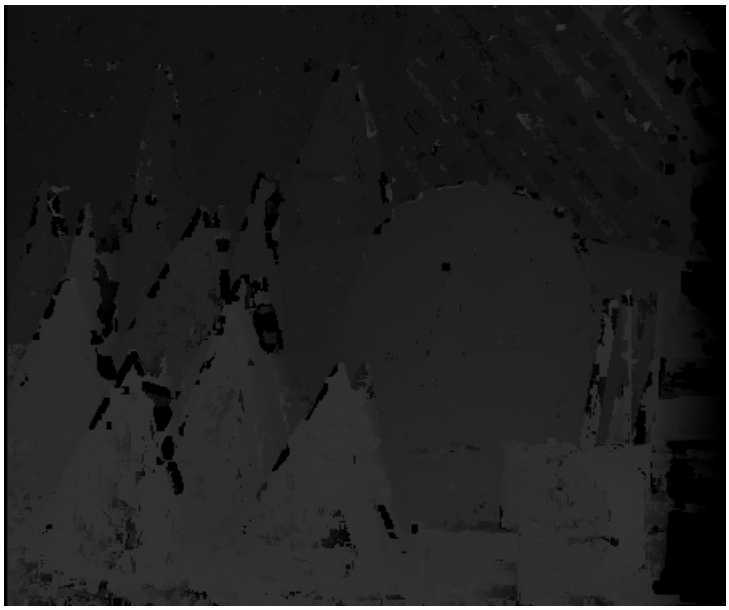
5.4 Cons

Les condicions per aquesta imatge són exactament les mateixes que en l'anterior.

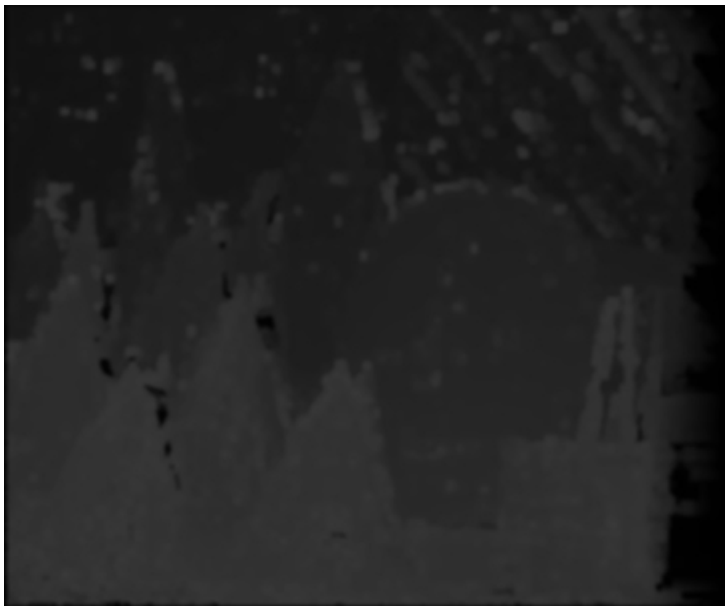
5.4.1 Imatges

Imatge esquerra original
Font: http://http://vision.middlebury.edu/stereo/

Imatge dreta original
Font: http://http://vision.middlebury.edu/stereo/



Mapa de disparitats calculat



Mapa de disparitats tractat i corregit

Nova imatge esquerra amb els forats tapats

Composició

Calen ulleres anaglifes per poder veure correctament la imatge

5.4.2 Temps de càlcul

Resolució de la imatge: 375×450

Càlcul del mapa de disparitats	20s (aprox.)
Tractament del mapa de disparitats	2.87s
Generació de la nova imatge	1.59s
Interpolació de forats	1.84s

Nota: El temps de càlcul de la tarja gràfica no s'ha pogut mesurar de manera precisa.

5.5 Porta

En aquest cas es tenen dos imatges de la mateixa escena, però ni les càmeres han estat orientades correctament, ni l'alçada ni la distància entre càmeres

és l'adequada. Cal, doncs, aplicar tot el procediment proposat en aquest projecte.

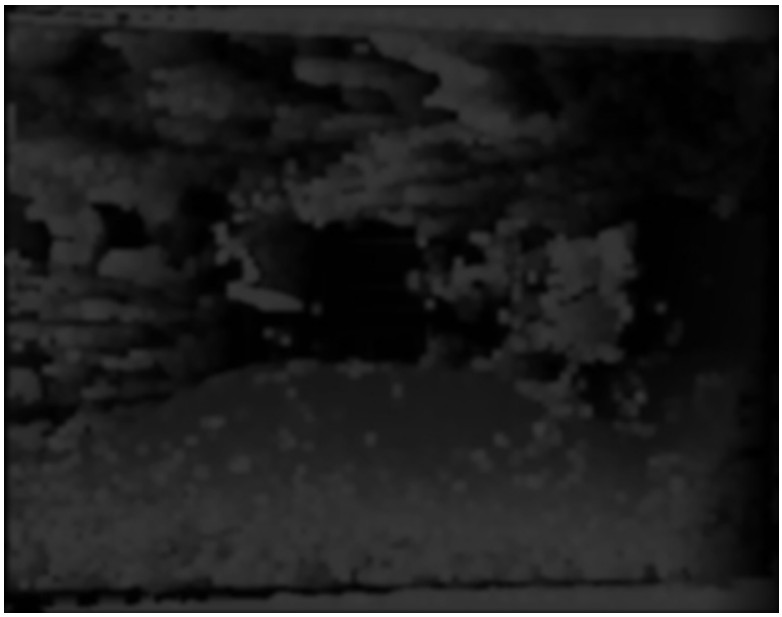
5.5.1 Imatges

Imatges esquerra i dreta originals
Font: http://http://profs.sci.univr.it/ fusiello/

Imatge esquerra rectificada

Imatge dreta rectificada

Mapa de disparitats calculat



Mapa de disparitats tractat i corregit

Nova imatge esquerra amb els forats tapats

Composició

Calen ulleres anaglifes per poder veure correctament la imatge

5.5.2 Temps de càlcul

Resolució de la imatge: 316×403

Detecció i identificació de punts característics (SURF)	11.09s
Rectificació	1.85s
Càlcul del mapa de disparitats	18s (aprox.)
Tractament del mapa de disparitats	1.97s
Generació de la nova imatge	1.23s
Interpolació de forats	1.85s

Nota: El temps de càlcul de la tarja gràfica no s'ha pogut mesurar de manera precisa.

Conclusions

En aquest projecte s'arriba a la conclusió de que, tot i que generar de manera artificial una imatge estereoscòpica a partir d'una posició qualsevol de les càmeres sigui molt més barat que obtenir el parell d'imatges a partir d'un set stèreo, la qualitat obtinguda no és comparable. No obstant, la sensació de profunditat sí que es pot aconseguir, per tant aquest procediment pot ser interessant per a aplicacions on no cal que la qualitat de la imatge sigui perfecta però que amb la profunditat n'hi hagi prou.

La dificultat principal és la obtenció del mapa de disparitats. El mètode proposat en aquest projecte és poc robust. El problema no és tant sols que els resultats poden ser poc acurats, sino que un mapa de disparitats erroni condiciona molt negativament tota la resta del procés.

En quant a la velocitat dels càlculs, l'avantatge és que molts dels procediments realitzats poden paral·lelitzar-se, no tant sols el càlcul de mapes de disparitat, sino també procediments com la interpolació de forats, o una part de la generació de la imatge esquerra virtual. Per tant, si es fa una paral·lelització gestionant adequadament la memòria de la tarja gràfica i la transferència de dades entre aquesta i la CPU, es poden arribar a aconseguir velocitats molt elevades.

Per tant, per millorar aquest procediment és essencial intentar obtenir un algoritme que permeti obtenir un mapa de disparitats el més correcte possible, i a la velocitat més elevada possible, altrament seria impensable la seva aplicació a procesos que requereixin un tractament a temps real.

Costos del projecte

Els costos d'aquest projecte es divideixen en 3 apartats: cost de l'equipament necessari per a la seva realització, cost de la programació i cost de funcionament.

Cost de l'equipament

- Ordinador amb procesador Athlon II, 2GB de RAM, 320GB de disc dur: 321€
- Tarja gràfica nVidia GeForce 9500GT: 96€
- Font d'alimentació 400W: 30€
- Webcam Soyntec (x2): 50€

Cost total de l'equipament: 497€

Cost de la realització

- Sou d'un programador becari: 6€/h
- Nombre d'hores estimades per a la programació de tots els algoritmes: 160h

Cost total de la realització: 960€

Cost de funcionament

- Potència consumida per la font d'alimentació de l'ordinador: 400W
- Preu del kWh per una potència contractada de 3.3kW (abril 2011): 0.140069€/kWh

Cost horari de funcionament: 0.0560276€/h

Costos totals

- Cost fix: 1457€
- Cost variable (per hora de funcionament): 0.0560276€/h

Estudi mediambiental

L'impacte ambiental de la posada en pràctica d'aquest projecte és baix. Els únics aspectes negatius són:

- El fet de precisar d'una tarja gràfica per poder realitzar càlculs en paral·lel implica un major consum d'energia elèctrica per part de l'ordinador. Cal disposar d'una font d'alimentació més potent que les que acostumen a venir de sèrie amb els ordinadors.
- Al final de la vida útil de l'ordinador, la tarja gràfica, les càmeres, etc, aquests es converteixen en residus que no poden llençar-se als contenidors normals del carrer. Cal portar-los a un punt verd o a una deixalleria.

A tall d'exemple, es mostra l'energia consumida durant una jornada laboral (8h) per un ordinador amb una font d'alimentació de $250W$ i un ordinador equipat amb una font d'alimentació de $400W$ com la usada en l'ordinador amb que s'ha realitzat aquest projecte:

Font $250W$	Font $400W$	Diferència
$2kWh$	$3.2kWh$	$1.2kWh$

Bibliografia

[1] L. Isara A. Fusiello. Quasi-euclidean uncalibrated epipolar rectification. 2008. [citat a la pàg. 28, 29]

[2] M. Stephens C. Harris. A combined corner and edge detector. *Proceedings of the Alvey Vision Conference, pàgines 147-151*, 1988. [citat a la pàg. 21]

[3] C. Fehn. Depth-image-based rengering (dibr), compression and transmission for a new approach on 3-d tv. *SPIE Conf. Stereoscopic Displays and Virtual Reality Systems, vol. 5291, pàgines 93-104*, 2004. [citat a la pàg. 41]

[4] T. Tuytelaars L. Van Gool H. Bay, A. Ess. Speed-up robust features (surf). *Computer Vision and Image Understanding (CVIU), vol. 110, no. 3, pàgines 346-359*, 2008. [citat a la pàg. 22, 24, 25, 26, 27]

[5] I. Barandiaran O Ruiz J. Congote, J. Barandiaran. Realtime dense stereo matching with dynamic programming in cuda. *CEIG'09, San Sebastián*, 2009. [citat a la pàg. 39]

[6] N-N. Zheng J. Sun. Stereo matching using belief propagation. *Transactions on pattern analysis and machine intelligence*, 2003. [citat a la pàg. 34]

[7] D. Wang L. Zhang, W. James. Stereoscopic image generation based on depth images. 2005. [citat a la pàg. 51, 52]

[8] W.J. Tam L. Zhang. Stereoscopic image generation based on depth images for 3d tv. *IEEE Transactions on broadcasting, vol. 51, pàgines 191-199*, 2005. [citat a la pàg. 51]

[9] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 2004.* [citat a la pàg. 22, 23]

[10] V. Caselles C. Ballester M. Bertalmio, G. Sapiro. Image inpainting. *SIGGRAPH 2000, pàgines 417-424*, 2000. [citat a la pàg. 48]

[11] R. McKenna Y. Chang M.M. Oliveira, B. Bowen. Fast digital image inpainting. 2001. [citat a la pàg. 48]

[12] D. Oram. Rectification for any epipolar geometry. [citat a la pàg. 28]

[13] Y-S Ho S-B. Lee. Discontinuity-adaptive depth map filtering for 3d view generation. 2009. [citat a la pàg. 51]

[14] J. Kosecká S.S. Sastry Y. Ma, S. Soatto. *An invitation to 3-D vision: from images to geometric models.* Springer, 2004. [citat a la pàg. 15]

Apèndix A

Conceptes matemàtics

En aquest capítol s'introduiran els conceptes matemàtics necessaris per a poder desenvolupar aquest projecte. Es parlarà de moviments rígids i es farà una breu introducció a la morfologia matemàtica.

A.1 Moviments rígids a l'espai

En aquesta secció es considera l'espai $\mathbf{R}^3$, i s'introduiran els moviments rígids, que són aplicacions que conserven la distància entre punts. Cal dir que, depenent del que convingui en cada moment, es tractarà $\mathbf{R}^3$ com un espai vectorial (els seus elements són vectors que es poden combinar linealment entre ells) o bé com un espai afí (els seus elements són punts, però la diferència entre dos punts és un vector).

A.1.1 Estructura mètrica de l'espai

Abans de res, cal introduir alguns conceptes importants que es faran servir, i que permetran dotar l'espai $\mathbf{R}^3$ d'altres estructures a part de la d'espai vectorial i la d'espai afí.

Definició 13. *El producte escalar euclidi de* $\boldsymbol{R}^3$ *és l'aplicació*

$$\begin{array}{rccc} <\cdot,\cdot>: & \boldsymbol{R}^3 \times \boldsymbol{R}^3 & \longrightarrow & \boldsymbol{R} \\ & (x,y) & \longmapsto & x_1y_1 + x_2y_2 + x_3y_3 \end{array}$$

on $(x_1, x_2, x_3), (y_1, y_2, y_3)$ *són les coordenades dels vectors* x, y *en la base ordinària.*

A partir de la definició es poden comprovar de manera immediata les següents propietats:

Proposició 1. *Propietats del producte escalar*

- *Simètric:* $< x, y >=< y, x >$
- *Lineal per l'esquerra:* $< \lambda x + \mu y, z >= \lambda < x, z > +\mu < y, z >$
- *Lineal per la dreta:* $< x, \lambda y + \mu z >= \lambda < x, y > +\mu < x, z >$
- *Definit positiu:* $< x, x >\geq 0$ $i < x, x >= 0 \Leftrightarrow x = 0$

Això permet introduir un concepte que serà molt útil per a poder tractar amb moviments rígids:

Definició 14. *Una base* (u_1, u_2, u_3) *de* $\boldsymbol{R}^3$ *es diu que és ortonormal si*

$$< u_i, u_j >= \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

A partir del producte escalar es pot definir la norma, mòdul o longitud d'un vector:

Definició 15. *La norma de* $\boldsymbol{R}^3$ *és l'aplicació donada per*

$$\begin{array}{rccc} \|\cdot\|: & \boldsymbol{R}^3 & \longrightarrow & [0,\infty) \subset \boldsymbol{R} \\ & x & \longmapsto & \sqrt{< x, x >} \end{array}$$

La norma compleix les següents propietats:

Proposició 2. *Propietats de la norma:*

- *Definida positiva:* $\|x\| \geq 0$ *i* $\|x\| = 0 \Leftrightarrow x = 0$
- $\|\lambda x\| = |\lambda|\|x\|$, *on* $\lambda \in \boldsymbol{R}$
- *Desigualtat triangular:* $\|x + y\| \leq \|x\| + \|y\|$

A més a més, a partir de les propietats de la norma i del producte escalar es pot deduir el següent resultat:

Proposició 3. *Desigualtat de Cauchy-Schwarz*

Siguin $x, y \in \boldsymbol{R}^3$, *aleshores:*

$$| < x, y > | \leq \|x\|\|y\|$$

La desigualtat de Cauchy-Schwarz és molt important i, entre moltes altres coses, permet assegurar l'existència d'un $\alpha \in [0, \pi]$ tal que $< x, y >= \|x\|\|y\| \cos \alpha$. A més a més, es pot comprovar que aquest α correspon a l'angle que formen els dos vectors, d'on es desprèn que dos vectors x, y són perpendiculars entre ells si, i només si, $< x, y >= 0$.

Finalment, s'acabarà aquest primer apartat definint una distància a $\mathbf{R}^3$:

Definició 16. *La distància de* $\boldsymbol{R}^3$ *és l'aplicació donada per*

$$\begin{array}{rccc} d: & \boldsymbol{R}^3 \times \boldsymbol{R}^3 & \longrightarrow & [0, \infty) \subset \boldsymbol{R} \\ & (x, y) & \longmapsto & \|x - y\| \end{array}$$

A partir de la definició i de les propietats de la norma, es dedueixen de manera immediata les següents propietats per a la distància:

Proposició 4. *Propietats de la distància*

- *Definida positiva:* $d(x, y) \geq 0$ *i* $d(x, y) = 0 \Leftrightarrow x = y$
- *Simètrica:* $d(x, y) = d(y, x)$
- *Desigualtat triangular:* $d(x, y) \leq d(x, z) + d(z, y)$

Amb aquesta aplicació es diu que $\mathbf{R}^3$ és un espai mètric, i és justament això el que calia per a poder parlar de moviments rígids a l'espai, tenir una manera de mesurar les distàncies entre punts.

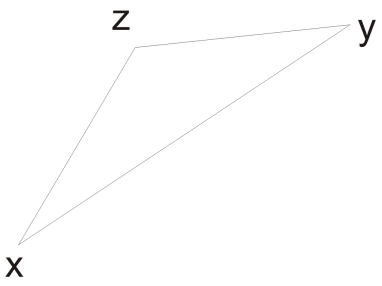


Figura A.1: Desigualtat triangular

A.1.2 Isometries lineals

En aquest apartat es parlarà dels moviments rígids lineals, també coneguts com isometries lineals, es classificaran els diferents tipus d'isometries, i s'aprofundirà una mica en l'estudi de les rotacions, que són el tipus d'isometria que es farà servir en el desenvolupament d'aquest projecte.

Definició 17. *Una isometria lineal de* $\boldsymbol{R}^3$ *és una aplicació* $f : \boldsymbol{R}^3 \to \boldsymbol{R}^3$ *tal que*

- f *és lineal:* $f(\lambda x + \mu y) = \lambda f(x) + \mu f(y)$
- f *és un moviment rígid:* $d(x, y) = d(f(x), f(y))$

La segona condició de la definició es pot reescriure de manera equivalent com $\|x\| = \|f(x)\|$, d'on es dedueix que si els valors propis de f tenen mòdul 1, i per tant $\det(f) = 1$ o $\det(f) = -1$.

Es té el següent resultat de molta utilitat per a comprovar si una aplicació és una isometria lineal:

Proposició 5. *(Caracterització de les isometries lineals)*

$f : \boldsymbol{R}^3 \longrightarrow \boldsymbol{R}^3$ *és una isometria lineal* $\Longleftrightarrow < x, y > = < f(x), f(y) >$

I d'aquest resultat se'n pot deduïr el següent, que estableix com són les matrius de les isometries lineals:

Proposició 6. *f és una isometria lineal si, i només si, la seva matriu A en una base ortonormal compleix que $A^tA = I_3$*

És a dir, si f és una isometria lineal serà invertible i si A és la seva matriu en una base ortonormal es complirà que $A^{-1} = A^t$.

Utilitzant aquest fet juntament amb que els valors propis de f tenen mòdul 1 es pot arribar a demostrar el següent resultat:

Proposició 7. *(Classificació de les isometries lineals de l'espai) Sigui $f : \boldsymbol{R}^3 \to \boldsymbol{R}^3$ una isometria lineal, aleshores:*

- $\det(f) = 1$ *si, i només si, existeix una base ortonormal de $\boldsymbol{R}^3$ tal que la matriu de f en aquesta base és*

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}$$

  En aquest cas, es diu que f és una rotació

- $\det(f) = -1$ *si, i només si, existeix una base ortonormal de $\boldsymbol{R}^3$ tal que la matriu de f en aquesta base és*

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}$$

En el cas en que f sigui una rotació, és a dir quan $\det(f) = 1$, la interpretació geomètrica és la següent: Si (u_1, u_2, u_3) és una base ortonormal respecte de la qual la matriu de f és la donada a la proposició, els vectors de la mateixa direcció que u_1 són fixos per f, de manera que no giren. El subespai generat per u_1 serà, doncs, l'eix de la rotació. Per altra banda, prenent un vector v que pertanyi al subespai engendrat per u_2 i u_3 aleshores l'angle format entre v i $f(v)$ és α, o sigui que α és l'angle de rotació. En el cas en que f no sigui una rotació, és a dir quan $\det(f) = -1$, f és una simetria respecte d'un pla seguida d'una rotació, però no s'aprofundirà més en aquesta interpretació perqué no és necessari per a aquest projecte.

De la interpretació anterior es dedueix el procediment a seguir per a obtenir les equacions d'una rotació f d'angle i eix coneguts:

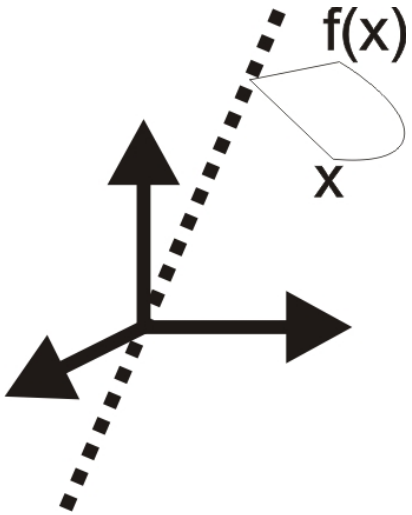


Figura A.2: Rotació lineal

- Construir una base ortonormal (u_1, u_2, u_3) prenent com a u_1 un vector en la direcció de l'eix, i construir la matriu de canvi de base S posant els vectors u_i com a columnes.
- Construir la matriu B de f en la base (u_1, u_2, u_3):

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}$$

- Calcular la matriu A de f en la base ordinària:

$$A = SBS^t$$

- Les equacions de f vindran donades per

$$f\begin{pmatrix} x \\ y \\ z \end{pmatrix} = A\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

A.1.3 Moviments rígids afins

Així com les isometries lineals són les aplicacions lineals que conserven la distància, els moviments rígids afins són aplicacions afins que conserven la distància:

Definició 18. *Un moviment rígid afí de* $\boldsymbol{R}^3$ *és una aplicació* $f : \boldsymbol{R}^3 \to \boldsymbol{R}^3$ *tal que*

- f *és una transformació afí*
- f *és un moviment rígid:* $d(x, y) = d(f(x), f(y))$

Tota transformació afí es descompon com una aplicació lineal seguida d'una traslació:

$$f(x) = \tilde{f}(x) + T$$

on $\tilde{f}$ és una aplicació lineal i T el vector de traslació. La propietat interessant és que l'aplicació $\tilde{f}$ és una isometria lineal, de manera que per a classificar moviments rígids lineals n'hi haurà prou amb classificar la isometria lineal associada. No s'entrarà en més detalls en la discusió de la classificació de moviments rígids afins, ja que per a aquest projecte n'hi haurà prou amb poder-los interpretar com una isometria lineal seguida d'una traslació

A.2 Morfologia matemàtica

En aquesta secció es farà una breu introducció a la morfologia matemàtica, i es veurà com es pot aplicar al tractament d'imatges. La morfologia matemàtica és una teoria inicialment concebuda per a tractar amb figures geomètriques, tot i que actualment s'ha extès i generalitzat. La referència principal per aquesta secció és Serra [**?**].

A.2.1 Reticles complets

L'estructura matemàtica bàsica per a desenvolupar la teoria de la morfologia matemàtica és el reticle complet. Abans de tractar amb aquest concepte, però, cal introduir-ne alguns de previs.

Definició 19. *Un conjunt* X *es diu que està parcialment ordenat si existeix una relació binària,* R*, que per a qualssevol* $a, b, c \in X$ *verifica les següents propietats:*

- *Reflexivitat:* aRa
- *Antisimetria:* $aRb,\ bRa \iff a = b$
- *Transitivitat:* $aRb,\ bRc \implies aRc$

Habitualment, se sol fer servir el símbol $\leq$ *per a denotar la relació* R*, i si* $a \leq b$ *es diu que* b *és més gran que* a*.*

Una observació important és que donats dos elements $a, b \in X$ no han d'estar necessàriament relacionats, i si no ho estan es diu que no són comparables. En el cas en que necessàriament es tingui que $a \leq b$ o $b \leq a$, per a tots els elements $a, b \in X$, aleshores es diu que el conjunt està totalment ordenat.

Exemple 3. *Exemples de conjunts parcialment ordenats:*

- $X = \boldsymbol{R}$ *amb* $\leq$ *sent la relació "menor o igual"habitual. De fet, amb aquest ordre,* $\boldsymbol{R}$ *és un conjunt totalment ordenat.*
- $X = \boldsymbol{R}^2$ *amb* $\leq$ *sent l'ordre lexicogràfic, és a dir,* $(x_1, x_2) \leq (y_1, y_2) \Leftrightarrow x_1 < y_1$ *o* $x_2 \leq y_2$ *si* $x_1 = y_1$*. Novament, el conjunt està totalment ordenat.*
- $X = \mathcal{P}(\boldsymbol{R})$ *(conjunt de parts dels nombres reals), sent* $\leq$ *la relació d'inclusió entre conjunts, és a dir:* $A \leq B \Leftrightarrow A \subset B$*. En aquest cas, el conjunt està parcialment ordenat però no totalment ordenat ja que hi ha elements, com* $\{1, 2\}$ *i* $\{1, 3, 4, 6\}$ *que no són comparables.*

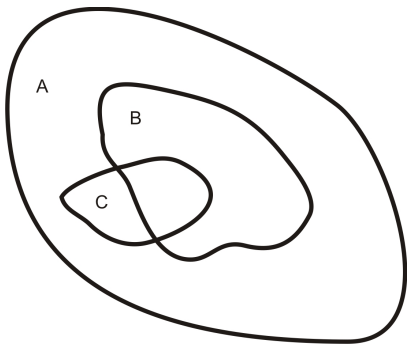


Figura A.3: $B \leq A$, $C \leq A$, però B i C no són comparables.

Definició 20. *Sigui X un conjunt parcialment ordenat. Sigui $Y \subset X$ un subconjunt. Un element $a \in X$ s'anomenarà cota superior de Y (resp. cota inferior de Y) si $b \leq a, \forall b \in Y$ (resp. $a \leq b, \forall b \in Y$).*

Es dirà que c és el suprem (resp. ínfim) de Y si és la menor de les cotes superiores (resp. la major de les cotes inferiors). El suprem i l'ínfim de Y es denoten per $\vee(Y)$ i $\wedge(Y)$ respectivament.

Cal observar que l'existència de cotes superiors i cotes inferiors no està assegurada. En cas que existeixin cotes superiors i/o cotes inferiors, tampoc està assegurada l'existència de suprems i/o ínfims, però quan aquests existeixen aleshores són únics:

Proposició 8. *Donat un subconjunt $Y \subset X$ d'un conjunt parcialment ordenat, en cas d'existir $\vee(Y)$ (resp. $\wedge(Y)$) aquest és únic.*

Exemple 4. *Prenent els conjunts parcialment ordenats de l'exemple anterior:*

- *Qualsevol nombre $a \geq 1$ és una cota superior del subconjunt $[0,1]$, i qualsevol nombre $b \leq 0$ és una cota inferior. El suprem és $\vee([0,1]) = 1$ i l'ínfim és $\wedge([0,1]) = 0$.*
- *Prenent el subconjunt $Y = \{(x,y) \in \boldsymbol{R}^2 | x \leq 1\}$, els elements $(a,b) \in \boldsymbol{R}^2$ tals que $a \geq 1$ són cotes superiors. No existeix el suprem de Y, ja que donada una cota superior sempre se'n pot trobar una de més petita. No existeixen cotes inferiors i per tant tampoc existeix l'ínfim de Y.*
- *Agafant $Y = \{(0, 1 + 1/n) | n \in \boldsymbol{N}\} \subset \mathcal{P}(\boldsymbol{R})$, els subconjunts donats pels intervals $(0,3), [-1,2)$, o el conjunt total $\boldsymbol{R}$, són algunes cotes superiors de Y, i els subconjunts $(0,1), (1/2, 1/2)$, o el conjunt buit $\emptyset$, són algunes cotes inferiors de Y. El suprem de Y és $\vee(Y) = \bigcup_{n \in \boldsymbol{N}} (0, 1+1/n) = (0,2)$ i l'ínfim és $\wedge(Y) = \bigcap_{n \in \boldsymbol{N}} (0, 1+1/n) = (0,1]$.*

Com ja s'ha comentat, i com es veu clarament en els exemples anteriors, no està assegurada l'existència de cotes superiors i inferiors i menys encara l'existència de suprems i ínfims. Al menys aquests últims en cas d'existir són únics, però dels exemples es desprèn que en cas d'existir el suprem i/o

l'ínfim d'un subconjunt $Y \subset X$ d'un conjunt parcialment ordenat, aquests no necessàriament seran elements del subconjunt Y. En el primer dels exemples es veu com tant el suprem com l'ínfim pertanyen al subconjunt. En el segon exemple no existeixen ni el suprem ni l'ínfim. En el tercer exemple el suprem pertany al subconjunt però l'ínfim no.

Definició 21. *Sigui X un conjunt parcialment ordenat. Si per a tot subconjunt $Y \subset X$ existeixen $\vee(Y)$ i $\wedge(Y)$ es diu que X és un reticle. Si, a més a més, $\vee(Y), \wedge(Y) \in Y$ aleshores X és un reticle complet*

A.2.2 Estructura de reticle complet en les imatges en escala de grisos

Per poder aplicar la morfologia matemàtica al tractament d'imatges caldrà dotar aquestes d'estructura de reticle complet. En el cas de les imatges en escala de grisos això és molt fàcil i intuitiu, en el cas de les imatges en color segueix sent senzill però no existeix una manera intuitiva de fer-ho. De totes formes, en aquest projecte no caldrà aplicar tècniques morfològiques a imatges en color, així que amb posar l'estructura de reticle complet a les imatges en escala de grisos serà suficient i, de fet, tenint en compte que l'objectiu és fer el tractament de les imatges amb ordinador, n'hi haurà prou amb posar l'estructura de reticle complet a les imatges digitals en escala de grisos.

Una imatge digital en escala de grisos era un parell (Ω, I) amb $\Omega = [1, N] \times [1, M] \subset \mathbf{R}^2$, $N, M \in \mathbf{Z}^+$, i $I : \Omega \to [0, 255] \cap \mathbf{Z}$ una aplicació que a cada píxel $(x, y) \in \Omega$ li assigna el seu valor de gris. Per a poder comparar dues imatges, és raonable imposar que el seu conjunt de píxels, Ω, sigui el mateix és a dir, dit d'una altra manera, que siguin imatges de la mateixa mida. Aleshores, donades dues imatges digitals en escala de grisos definides sobre el mateix conjunt de píxels, $(\Omega, I) i (\Omega, I')$, l'ordre vindrà donat per la relació

$$(\Omega, I) \leq (\Omega, I') \Longleftrightarrow I(x, y) \leq I'(x, y), \forall (x, y) \in \Omega$$

Proposició 9. *Fixat un conjunt $\Omega = [1, N] \times [1, M] \subset \boldsymbol{Z}^2$, la relació anterior defineix un ordre parcial sobre el conjunt de totes les imatges digitals en escala de grisos que tinguin Ω com a conjunt de píxels.*

Si el conjunt de píxels Ω es pot sobreentendre s'escriurà directament $I \leq I'$. Hi haurà imatges que no seran comparables.

Exemple 5. *Sigui* $\Omega = [1,3] \times [1,3] \subset \boldsymbol{Z}^2$. *Es consideren les imatges*

$$I_1 = \begin{bmatrix} 0 & 100 & 255 \\ 3 & 200 & 240 \\ 200 & 200 & 220 \end{bmatrix}$$

$$I_2 = \begin{bmatrix} 100 & 100 & 255 \\ 100 & 200 & 255 \\ 200 & 200 & 255 \end{bmatrix}$$

$$I_3 = \begin{bmatrix} 50 & 100 & 255 \\ 200 & 200 & 255 \\ 200 & 200 & 255 \end{bmatrix}$$

Aleshores $I_1 \leq I_2$, $I_1 \leq I_3$, *i* I_2 *i* I_3 *no són comparables.*

Exemple 6. *De les imatges següents, la primera és menor o igual que la segona:*

Figura A.4: Ordre en imatges

Amb aquest ordre s'aconsegueix l'estructura buscada de reticle complet per a les imatges digitals en escala de grisos.

Proposició 10. *Fixat un conjunt* $\Omega = [1,N] \times [1,M] \subset \boldsymbol{Z}^2$, *el conjunt de totes les imatges digitals en escala de grisos sobre* Ω *és un reticle complet amb la relació d'ordre definida anteriorment.*

A.2.3 Erosions i dilatacions

Un cop presentada l'estructura de reticle complet i havent dotat d'aquesta a les imatges digitals en escala de grisos, es definiran les dos transformacions elementals, que seran de gran importància i utilitat en el tractament d'imatges: les erosions i les dilatacions.

Definició 22. *(Imprecisa) Sigui X un reticle complet.Una aplicació $B : X \to \mathcal{P}(X)$ és un element estructurant si per a qualssevol $x \in X$, $x \in B(x)$ i $B(y)$ té la mateixa forma que $B(x)$ per a tot $y \in X$*.*

La idea és que a cada element de X se li assigna un veïnat (d'aquí el fet que $x \in B(x)$) que té la mateixa forma per a qualsevol punt. Per exemple, en subconjunts del pla, un element estructurant pot ser centrar a cada punt un cercle d'un determinat radi constant, o un centrar a cada punt un quadrat de mides constants,...

Definició 23. *Sigui X un reticle complet i $B : X \to \mathcal{P}(X)$ un element estructurant.*

- *Una erosió per B és una aplicació $\varepsilon : X \to X$ tal que $\varepsilon(x) = \wedge(B(x))$.*
- *Una dilatació per B és una aplicació $\delta : X \to X$ tal que $\delta(x) = \vee(B(x))$.*

Hi ha moltes més transformacions morfològiques, però aquestes dos són les bàsiques i, de fet, les úniques que es faran servir en aquest projecte.

A.2.4 Erosions i dilatacions d'imatges en escala de grisos

Recordant l'estructura de reticle complet amb que s'ha dotat a les imatges digitals en escala de grisos definides sobre un determinat conjunt de píxels $\Omega = [1, N] \times [1, M] \subset \mathbf{Z}^2$, es procedirà a definir les erosions i dilatacions més habituals.

*Aquesta definició que s'ha donat és molt imprecisa. Per a fer-ho de manera rigorosa primer s'hauria de dotar X d'estructura topològica i aleshores dir que $B(x)$ i $B(y)$ són *homeomorfs* o *topològicament equivalents*

Primer de tot, cal definir quin serà l'element estructurant. En totes les aplicacions de la morfologia usades en aquest projecte s'ha fet servir com a element estructurant l'aplicació que a cada píxel $(x, y) \in \Omega$ li assigna el subconjunt format per la finestra de mida $n \times n$ (amb n imparell) centrada en (x, y).

L'erosió d'una imatge digital en escala de grisos (Ω, I) per l'element estructurant $B(x, y)$ (finestra $n \times n$ centrada en (x, y)) és una nova imatge $(\Omega, \tilde{I})$ tal que

$$\tilde{I}(x, y) = \min_{(i,j) \in B(x,y)} I(i, j)$$

És a dir, a cada píxel (x, y) se centra una finestra de mida $n \times n$, es mesuren totes les intensitats de gris dels píxels de la finestra, i d'aquestes s'assigna a $I(x, y)$ la de valor mínim.

La dilatació d'una imatge digital en escala de grisos (Ω, I) per l'element estructurant $B(x, y)$ (finestra $n \times n$ centrada en (x, y)) és una nova imatge $(\Omega, \tilde{I})$ tal que

$$\tilde{I}(x, y) = \max_{(i,j) \in B(x,y)} I(i, j)$$

La idea de la dilatació és exactament la mateixa però prenent la intensitat de gris màxima de la finestra en comptes de la mínima.

Exemple 7. *Erosions i dilatacions de la primera imatge de l'exemple 6:*

Figura A.5: Erosió i dilatació amb finestra de mida $n = 3$

Com es pot observar a les imatges de l'exemple, hi ha una banda a la frontera de la imatge que queda sense tractar, la mida de la qual va creixent a mesura que s'augmenten les dimensions de l'element estructurant. Aquesta banda s'anomena banda de brossa, i són píxels on no es pot centrar l'element estructurant perquè aquest se surt de la imatge.

Figura A.6: Erosió i dilatació amb finestra de mida $n = 5$

Figura A.7: Erosió i dilatació amb finestra de mida $n = 7$

Exemple 8. *(Aplicació de l'erosió a la detecció de contorns)*

Una aplicació de l'erosió és la detecció de contorns. Tot i que existeixen algoritmes molt més sofisticats, com els de Canny i Sobel, amb l'erosió es pot aconseguir un detector de contorns molt senzill. Si (Ω, I) *és la imatge original i* $(\Omega, \tilde{I})$ *la imatge erosionada, aleshores* $(\Omega, I - \tilde{I})$ *serà la imatge amb els contorns.*

A.3 Convolució

La convolució és una eina de gran importància en el tractament d'imtages. Primer es veuran les definicions i propietats generals de la convolució, i a continuació es veuran algunes aplicacions al tractament digital d'imatges.

Figura A.8: Detecció de contorns amb finestra de mida $n = 5$

A.3.1 Definicions i propietats

Definició 24. *(Convolució de senyals continus)*

Siguin $f : \boldsymbol{R}^n \to \boldsymbol{R}$ *i* $g : \boldsymbol{R}^n \to \boldsymbol{R}$ *funcions de quadrat integrable, és a dir* $\int_{\boldsymbol{R}^n}(f(x))^2 dx < \infty$ *(i anàlogament per* g*). La convolució de* f *per* g *és la funció definida per*

$$(f * g)(x) = \int_{\boldsymbol{R}^n} f(y)g(x-y)dy$$

La funció g *s'anomena nucli de la convolució*

Definició 25. *(Convolució de senyals discrets)*

Siguin $f : \boldsymbol{Z}^n \to \boldsymbol{R}$ *i* $g : \boldsymbol{Z}^n \to \boldsymbol{R}$ *dues funcions de quadrat sumable sobre* W*, és a dir* $\sum(f(k))^2 < \infty$ *(i anàlogament per* g*). La convolució de* f *per* g *és la funció definida per*

$$(f * g)(k) = \sum_{m \in \boldsymbol{Z}^n} f(m)g(k-m)$$

La funció g *s'anomena nucli de la convolució*

En ambdós casos es verifica la següent llista de propietats:

Proposició 11. *Propietats de la convolució*

- *Bilinealitat:*
  - $(\lambda f + \mu g) * h = \lambda f * h + \mu g * h$
  - $f * (\lambda g + \mu h) = \lambda f * g + \mu f * h$
- *Commutativitat:* $f * g = g * f$
- *Associativitat:* $(f * g) * h = f * (g * h)$

De la bilinealitat es desprèn que es compleix també la propietat distributiva. Cal observar que no existeix un element neutre per a la convolució.

Tot i que en aquest projecte, lògicament, només s'aplicarà la convolució al tractament d'imatges, es tracta d'un recurs matemàtic de molta utilitat en molts camps de la ciència i l'enginyeria: electrònica, tractament de so, estadística, òptica,...

A.3.2 Convolució aplicada al tractament d'imatges

Donada una imatge digital en escala de grisos (Ω, I), la seva convolució pel nucli g és la imatge $(\Omega, I * g)$. Triant adequadament el nucli es poden aconseguir diversos efectes sobre la imatge. Si la imatge és en color es fa la convolució dels tres canals de color per separat, és a dir si $I = (I_R, I_G, I_B)$ aleshores $I * g = (I_R * g, I_G * g, I_B * g)$.

Seguidament es presenten algunes aplicacions de la convolució d'imatges. Per a fer els exemples s'ha pres la primera imatge de l'exemple 6.

Accentuació

- Nucli:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 5 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- Exemple:

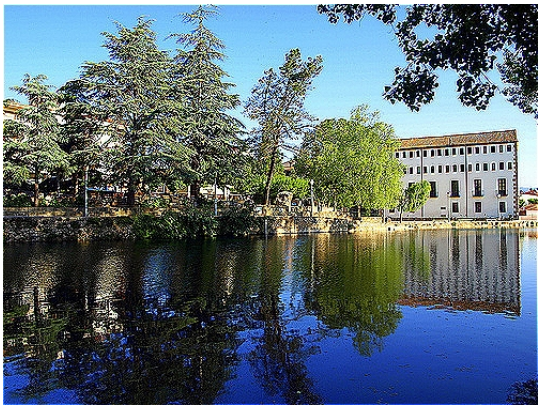


Figura A.9: Accentuació

Difuminat

- Nucli:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- Exemple:

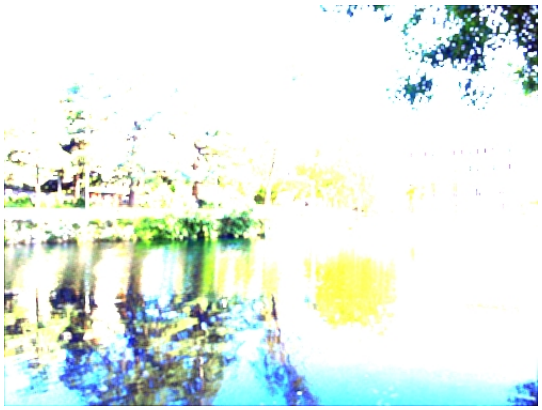


Figura A.10: Difuminat

Detecció d'eixos

Es converteix abans la imatge a escala de grisos, per a obtenir millors resultats.

- Nucli:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

- Exemple:

Figura A.11: Detecció d'eixos

Realçament

- Nucli:

$$\begin{bmatrix} -2 & -1 & 0 \\ -1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

- Exemple:

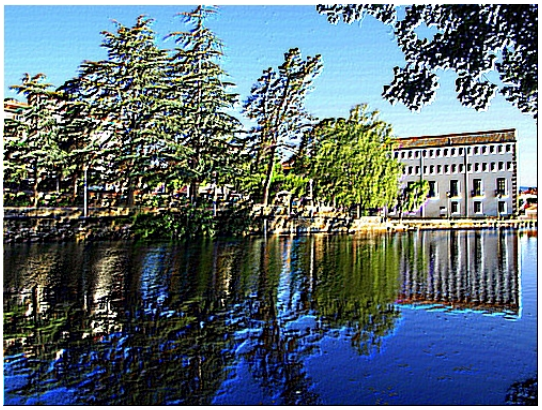


Figura A.12: Realçament